FOUNDED 1925
INCORPORATED BY
ROYAL CHARTER 1961

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

THE RADIO AND ELECTRONIC ENGINEER

The Journal of the Institution of Electronic and Radio Engineers

VOLUME 36 No. 5

NOVEMBER 1968

Co-operation in Space Research and Communications

THIS issue of *The Radio and Electronic Engineer* contains three papers dealing with different aspects of space science and technology: Dr. Maurice Ponte's Clerk Maxwell Memorial Lecture takes as its theme 'Spatial Communications' and discusses communication satellites and radio astronomy; Mr. L. H. Bedford assesses the potentialities of satellites as television broadcasting stations; the paper by Mr. D. Weighton, presented at the Symposium on the *Ariel III* satellite, describes the receiver of a galactic noise research experiment. These three papers thus do not overlap to any appreciable extent and it would be possible to contemplate further independent papers on radio and electronic applications in space; for instance, the problems of attitude control and power supplies for artificial satellites, navigational and meteorological satellites, telemetry and control for deep-space probes, to name but a few.

The variety of techniques and applications is a measure of the amazing burgeoning of this new field for man's endeavours, for, as we have pointed out recently†, it is only 11 years since the first artificial Earth satellites were put into orbit. During recent weeks there have been reported such achievements as the protracted flight of the *Apollo VII* and its three-man crew, the Soviet *Zond 5* circumlunar operation and the *Soyuz 2* and *3* flights; and, on a much more modest scale, though significant as a co-operative venture, the successful launching of the European ionospheric research satellite *ESRO I*, now known as *Aurorae*: all these projects are basically concerned with pure or applied research. The growing networks of communication satellites have now reached the degree of flexibility at which the failure to go into orbit of *Intelsat III* could be made good as a temporary measure by using the already orbiting *ATS 3* (Applications Technology Satellite) and using (in the United Kingdom) the Goonhilly 1 station instead of the new Goonhilly 2 as had been planned. The effectiveness of these measures was seen all over the world in the transmissions of television pictures from the Olympic Games in Mexico City. Another network of communication satellites is the *Molniya* eight-hour orbit series which is used for relaying television broadcasts across the Soviet Union.

The *Intelsat* system of satellites is, in effect, a commercial venture by some 60 countries and there is considerable complementary activity in many countries in setting up Earth stations to operate inter-continental links. Cable and Wireless Ltd., for instance, are setting up stations in the West Indies, East Africa, Bahrain and Hong Kong in association with local telecommunications authorities. Co-operation is also a notable feature of all aspects of the Intelsat scheme as many large European electronics companies are assembling satellites and supplying other equipment for the *Intelsat IV* series planned to be launched in 1970.

The pace of advance in 'radio techniques and space research' can fairly be said to have more than met the expectations of the papers read at the Institution's Convention on that theme held in 1961. The emphasis given in the Sixth Clerk Maxwell Lecture to the development in spatial communications undoubtedly heralds still further progress in this new branch of technology in which both European and international co-operation is essential.

F. W. S.

† 'Astronomy, space and radio', *The Radio and Electronic Engineer*, **36**, No. 2, p. 69, August 1968.

INSTITUTION NOTICES

New Zealand Advisory Council

The Council of the Institution has set up a New Zealand Advisory Council which will initiate and co-ordinate meetings of members, in addition to acting as a central organization to ensure the participation of the I.E.R.E. and its members in all conferences and other activities which interest or affect New Zealand electronic radio engineers. The first chairman of the Advisory Council will be Professor Leslie Kay, Ph.D., C.Eng. (Fellow), of the University of Canterbury, Christchurch.

The Council will be responsible for issuing *Proceedings* of the New Zealand Division which will have the object of keeping members in touch with their colleagues and with the general activities of the Institution. Papers and other articles or notes of interest are invited for the *N.Z. Proceedings* and these should be sent to Mr. C. W. Salmon, Honorary Editor, at P.O. Box 3381, C.P.O., Auckland, New Zealand. Mr. Salmon is also to act as Honorary Treasurer of the Advisory Council. Mrs. E. M. Keating has been appointed secretary and correspondence in connection with New Zealand affairs should be sent to her at the Office of the New Zealand Advisory Council of the I.E.R.E., c/o Department of Electrical Engineering, University of Canterbury, Private Bag, Christchurch.

Gandhi Centenary

The Director of the Institution, Mr. Graham D. Clifford, has accepted an invitation to be a member of the United Kingdom National Committee for the Gandhi Centenary.

Conference on Digital Methods of Measurement

The I.E.R.E. is organizing, with the co-sponsorship of the Institution of Electrical Engineers and the Institute of Electrical and Electronics Engineers (U.K. and Republic of Ireland Section), a Conference on Digital Methods of Measurement to be held at the University of Kent at Canterbury, from Wednesday, 23rd to Friday, 25th July 1969.

The Joint Organizing Committee, which is under the chairmanship of Mr. K. J. Dean (Fellow), has defined the scope of the Conference as follows:

Measurement of physical quantities;

Counters, digital voltmeters and data loggers;

Programmed instruments;

Computer-controlled instruments;

Measurement problems in digital communication systems including p.c.m., data and telemetry.

Offers of papers, which should be accompanied by synopses of 200 to 300 words, are invited; these should be sent to the Secretary of the Organizing Committee for the Conference on Digital Methods of Measurement, I.E.R.E., 8-9 Bedford Square, London, W.C.1, as soon as possible.

The headquarters for the Conference will be Eliot College (the first to be established of the three colleges which now make up this new University). The technical sessions of the Conference will be held in the lecture theatre of the Cornwallis Building. Further information will be published in the *Journal* as soon as it is available.

Electronic Weighing Conference

A two-day conference on Electronic Weighing, organized by the Institution, was held in London last month. The final Conference Proceedings, which will contain fourteen papers and reports of the discussions, will be published early in 1969, price £3 10s. per copy. A few sets of preprints of papers are still available, price £2 10s. each.

Orders should be sent to the Publications Department, I.E.R.E., 8-9 Bedford Square, London, W.C.1.

Convention on Modernizing Education and Training

An international convention (COMET 69) will be held at Grosvenor House, London, from 2nd to 6th September 1969, to survey the application, development, techniques and evaluation of systems and aids in the field of education and training.

The convention will consist of a conference, exhibition and demonstrations of audio-visual aids and systems being used in teaching situations.

The conference will be organized by the Institution of Electrical Engineers and the National Council for Educational Technology in collaboration with the Institution of Electronic and Radio Engineers, the Institute of Mathematics and its Applications, the Royal Television Society and the Institute of Electrical and Electronics Engineers (U.K. and Republic of Ireland Section). The associated exhibition will be organized by the Industrial Council for Educational and Training Technology. The I.E.R.E. is represented on the convention Management Committee by Mr. M. H. Evans (Member).

Further details may be obtained from the Conference Department, I.E.E., Savoy Place, London, W.C.2.

The Sixth Clerk Maxwell Memorial Lecture

SPATIAL COMMUNICATIONS

by

Maurice J. H. Ponte, D.Sc., Ac.Sciences, Paris.

Delivered at the Institution's Convention in the University of Cambridge, on 3rd July 1968

1. Introduction

May I say how deeply I appreciate the honour of having been asked to deliver this year's Clerk Maxwell Memorial Lecture in this most distinguished place, which is a link between one of the scientists who transformed the world and one of the most distinguished universities ever to have existed—Cambridge—whose name is indissolubly linked with each step forward in Science.

For me, your invitation represents a kind of crowning event, for I have been very lucky in my relations with your country. Indeed, I was fortunate enough to work at the Royal Institution with Sir William Bragg, in the Davy–Faraday Laboratory still so full of reminders of those great scientists and of Faraday's fundamental discovery of induction in the year that Maxwell was born. The collection of modest apparatus and notebooks showed that Faraday was the typical English experimental physicist who liked to use his hands and who allowed himself to be guided by experiment without any preconceived ideas. I should like to tell you a little story, a true one, about my first meeting with Sir William Bragg, a long time ago now, at the end of 1925. Imbued with mathematical studies on the structure of crystals, to which the great physicist had made such a big contribution, I wanted to study the effects of imposing electrical fields on crystal lattices and of inserting foreign ions. Sir William listened to me in his usual good-natured way, encouraged me, then said: 'Go and see my assistant, Shearer, he'll give you whatever you need.' I was given a large piece of brass, a lamp-glass, a kettle with which to make an electrolytic interrupter, some wire, some sheet-iron with which to make a transformer, and various other bits and pieces to complete the X-ray installation. The interrupters had the annoying habit of exploding from time to time and their batteries were up on the roof. Whenever we heard the dull sound of an explosion, each of us would look at his apparatus and the unlucky one would proceed upstairs to repair the damage. Looking at all the humble apparatus with which Faraday had brilliantly reconstructed or tried to reconstruct the unit of nature's forces, we were filled with hope, determination and faith in our own equipment.

And now I have been given an opportunity to share in the tribute you pay here periodically to Maxwell, that great admirer of Faraday, to whom he was the perfect complement, thanks to a mathematical culture which Faraday, with his modest background, had been unable to acquire. It is another wonderful occasion for me, therefore, to be back here in Cambridge, where I learnt so much in 1926 and where I recall how proud I was to be wearing the Cambridge light blue at the Boat Race that year. So may I extend my heartfelt thanks to you once more?

Nothing has been left unsaid about Maxwell, but I should like to stress three points which seem to me fundamental, one of which led to my choice of subject.

One is filled with awe at the thought of speaking in a place where Clerk Maxwell taught. Maxwell was an outstanding teacher, though his premature death prevented him from personally completing his didactic work. Thus, Sir J. J. Thomson, another of Cambridge's many brilliant products, wrote in 1891: 'I have attempted to verify the results which Maxwell gives without proof: I have not in all instances succeeded. . . .' In his didactic work, one of Maxwell's virtues was his conscientiousness. It was in preparing papers on Faraday's lines of force, to be read before the Cambridge Philosophical Society, that he began to construct the mathematical theory of electromagnetic fields. He also averred that nothing could be more salutary than to re-read original memoirs, which he did copiously in the case of Faraday and which he strongly recommended to students: 'It is of great advantage to the student of any subject to read the original memoirs on that subject, for science is always most completely assimilated when it is in the nascent state'. (Preface to 'Treatise on Electricity and Magnetism'.)

At the same time, Maxwell was a precursor in spheres other than pure science. He realized the value of applied research and the consequences on a nation's economy, which was most meritorious in 1873 for a scientist. To quote Maxwell: '. . . The important applications of electromagnetism to telegraphy have also reacted on pure science by giving a commercial value to accurate electrical measurements and by affording to electricians the use of apparatus on a scale which greatly transcends that of any ordinary laboratory. The consequences of this demand for electrical knowledge, and of those experimental opportunities for acquiring it, have been already very great, both in stimulating the energies of advanced electricians, and in diffusing among practical men a degree of accurate knowledge which is likely to conduce to the general scientific progress of the whole engineering profession'. These words, written in 1873, clearly define the initial role of Science, at a time when the terms 'basic research', 'applied research', 'technology' and 'development' had still to be coined. If one adds to this the fact that Maxwell produced a splendid synthesis of the work done on electricity and magnetism in Europe in fifty years—one of his equations may be said to symbolize Ampère's work and the other Faraday's—then he can be said to have demonstrated the universality of science at a time when politics had not yet complicated the problem.

Lastly, it must be remembered that Maxwell was the first to construct a universally true theory placed at man's disposal. True, Newton—another of the Cambridge 'greats'—propounded the laws of gravitation that govern the motions of all bodies in the Universe, but so far, at any rate, man has not been able to do anything about it: he observes the facts and determines the laws. With the theories of electromagnetism, completed by Lorentz, and of relativity, man has acquired the ability, still on a modest scale it is true, to explore the Universe more and more exhaustively and for the first time influence it. All these reasons made me choose 'Spatial Communications' as my subject here. This subject provided an illustration of Maxwellian practical and scientific applications in the Universe and calls for international, or at any rate European, co-operation, and it is my ardent hope that the present difficulties encountered will be of a temporary nature only. Spatial communications provide a tremendous drive for technology and research as well as for teaching: in a word, it is a typically Maxwellian subject. My lecture will be a survey paper: I shall not teach anything to the specialists but it is sometimes necessary to summarize the fast evolving questions and examine where we are going.

It can be divided into two parts:

'Man-made' communications, which include the commands to spacecraft that enable them to do their job of exploring the world, and the communications used to transmit information from one point on the planet to another. (But not direct television communications from satellites, a controversial subject. so well studied by the paper of Mr L. H. Bedford given in the last C.I.T. session.†)

'Natural' communications, which originate from the universe and are picked up by radio observatories, or what a poet like Maxwell might have called 'the voice from the dark heavens'.

2. 'Man-made' Spatial Communications

2.1. *General Principles: Balance-sheet of the Link*

2.1.1 Evaluation of the power received over the r.f. carrier

Referring to Fig. 1, the transmitter (of power P_1) drives a suitably-orientated directive antenna (of gain G_1). The energy is propagated through space over a total distance *d*. Certain additional losses, due to the media traversed, may add themselves to the theoretical attenuation of propagation through empty space. Part of the energy is picked up at the reception point behind the appropriately-trained directive antenna (of gain G_2).

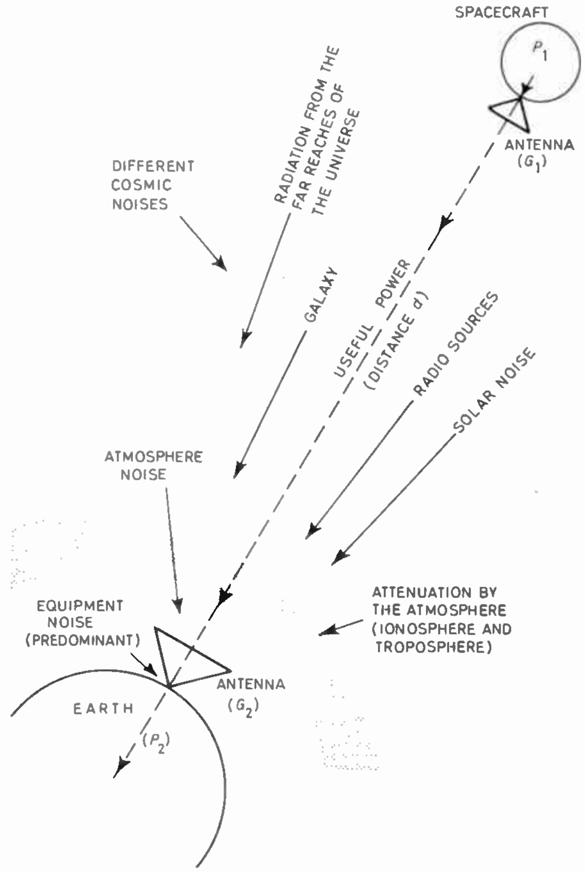


Fig. 1. Diagram of a radio frequency link between Earth and satellite.

† Bedford, L. H., 'Television from satellites', *The Radio and Electronic Engineer*, 36, No. 5, pp. 273–83, November 1968.

The ratio between the power received and the power transmitted, which characterizes the energy efficiency of the link is expressed in decibels by

$$10 \log_{10} \frac{P_2}{P_1} = G_1 + G_2 - L_S - L_D \qquad \ldots\ldots(1)$$

L_S represents the loss due to propagation through empty space irrespective of the directive effect of the antennae. This quantity, which can provide the basis for reckoning tables, depends both on the distance d (km) and the frequency f (MHz), in accordance with the classic relations:

$$L_S = 32{\cdot}45 + 20 \log_{10} f + 20 \log_{10} d \qquad \ldots\ldots(2)$$

Let L_D represent other losses, including in particular:

(a) additional attenuations due to the propagation media (chiefly the Earth's atmosphere, including the troposphere and the ionosphere); if f lies between 50 MHz and 10 000 MHz (the interval constituting the space 'window'), one can disregard any significant absorption effects;

(b) certain safety margins, to allow for accidental drops in performance in the equipment.

Expressing P_1 and P_2 in dB with respect to a level of 1 W (dBW), we may write:

$$P_2 = P_1 + G_1 + G_2 - L_S - L_D \qquad \ldots\ldots(3)$$

Let us examine the orders of magnitude involved. The antenna on the spacecraft may have a diameter of the order of 1 metre (which, assuming 2000 MHz, gives a gain of about 25 dB). Feasible diameters on the ground, on the other hand, can be much greater, say $D = 25$ m, giving a gain of about 50 dB for 2000 MHz.

The losses L_S are considerable, being in the region of between 250 and 295 dB for planetary exploration in the solar system. The big difference between this figure and the loss involved in links utilizing Earth satellites as intermediate relays for terrestrial telecommunications will be noted. In the latter case the loss is only about 155 dB. The difference of 140 dB (or 10^{14}) is considerable. Some consolation may be gained from the thought that in the case of the star Alpha Centauri, L_S would attain 372 decibels.

2.1.2. Reception noise

The quality of the link depends on the ratio between the power received P_2 and the noise power P_K measured at the receiving device's input.

In the most common case of noise of uniform density, the powers of the different causes of noise can be expressed as:

$$P_K = 1{\cdot}37 \times 10^{-23}\, T_K B \text{ erg/second} \qquad \ldots\ldots(4)$$

where B is the bandwidth considered (in hertz) and T_K the equivalent noise temperature (in degrees Kelvin).

When many noises of different non-correlated origins add together in the same circuit, the noise powers or equivalent temperatures are added.

Taking the most typical case for the ground telemetry receiving station, two broad categories of noise can be distinguished:

(a) noise due to imperfections in the apparatus;

(b) noise picked up from the outside medium.

2.1.3. Noise from apparatus

This category mainly includes the following:

(i) The noise temperature corresponding to the natural noise of the receiver's circuits depends on the quality of the input devices and will therefore vary greatly according to their nature:

circuits employing classic vacuum tubes:

$T_K = 2000°$ K approx.

parametric amplifier: $T_K = 100°$ approx.

maser input: T_K between 5 and 20° K.

(ii) The noise temperature relating to the waveguide (or to the coaxial cable) connecting the antenna to the receiver has the order of magnitude for T_K between 15 and 25° K.

(iii) The noise temperature stemming from the thermal radiation from the ground surrounding the antenna, is picked up by the latter via its secondary lobes (due to imperfections in the receiving antenna).

The cumulative effect of (i), (ii) and (iii) above, in the case of the most advanced type of ground station utilizing a maser device, is an overall apparatus temperature between 20 and 45° K.

In the reverse case (ground to spacecraft), one cannot lavish all the care one would like on the receiver's input circuitry, owing to small size requirement and the lower available electrical power. The noise temperature of the input circuits (using conventional semiconductor) is then of the order of 1800° K and thus very much greater than the ground station receiver.

2.1.4. Noise from the surrounding medium picked up by the antenna

This may be of cosmic origin, or of terrestrial origin (due to the atmosphere).

(1) 'Cosmic' noise includes:

the radiation due, at the very least, to the 'background' radiation from the Universe, which is very low but not negligible,

radiation from the region of the Galaxy, radiation picked up from specific 'radio sources',

solar radiation, chiefly accidental, picked up by the antenna's secondary lobes.

(2) 'Terrestrial' noise originating from the atmosphere surrounding the Earth, consisting chiefly of radiation from water-vapour and oxygen molecules.

In any event, save for the very exceptional cases due to incorrectly directed antennae, the temperatures associated with noise picked up are on the whole negligible at the present stage of the art in comparison with the noise temperature due to the equipment itself, and this is true even in the case of the most advanced type of ground receiving station.

2.2. *Scientific and Exploration Missions*

Scientific interplanetary exploration missions performed with spacecraft call for a comprehensive telecommunications system so as to make it possible:

(a) to keep the spacecraft's progress under constant surveillance, i.e. information relating to navigation,

(b) to transmit command from Earth to the spacecraft, i.e. command guidance signals, and

(c) to receive telemetry data from the spacecraft, relating either to scientific data picked up aboard it in fulfilment of its mission, or to information relating to operation of the spacecraft itself (engineering telemetry).

For obvious reasons of economy, the radio link established (from ground to spacecraft and from spacecraft to ground) must handle all three categories of information.

The information to be transmitted or received will be characterized either by its nature or its code. We therefore have a sequence of digits (0 or 1) succeeding one another at a uniform rate, which must be transmitted by radio, i.e. by the usual process of modulating a 'carrier' (or, in view of the high final frequencies used, by modulating an intermediate 'sub-carrier').

The procedure known as p.s.k. (phase shift keying) is tending to become the standard practice adopted because of its high performance capabilities in the presence of other signals. This type of modulation consists in considering two possible states of the sub-carrier phase, for example the phase 0 for the digit 0 and the phase π for the digit 1, or vice versa. To enable these phase permutations to be detected, the phase reference of origin must be available at the reception end, implying the need for a 'synchronous detection' device, which is one of the principal elements of the transmission system.

Some numerical data will serve to fix ideas on a more concrete basis. The figures relate to U.S. space probe programmes, about which published material is most readily available, i.e. to the *Pioneer* (solar exploration) and *Mariner* (missions to Mars and Venus).

P.c.m./p.s.k./ph.m. modulation is used. The first abbreviation (p.c.m. = pulse code modulation) indicates that the data are coded and multiplexed digitally, in other words take the form of a succession of digits.

The p.s.k. term denotes the 'sub-carrier' which must be transmitted by the high frequency, or radio frequency, ultimately radiated. This frequency is modulated by the sub-carrier, using classic phase modulation, which accounts for the third abbreviation ph.m.

(a) Final radio frequency (r.f.) used for transmission.

This is now around 2200 MHz, following an initial period when lower frequencies of between 400 and 1000 MHz were used. To avoid interference, the Earth-to-spacecraft and spacecraft-to-Earth transmission channels use slightly off-set frequencies.

(b) Sub-carrier (p.s.k.-modulated)

This employs frequencies of around 2000 Hz, 600 Hz or 150 Hz, depending on the digit transmission rate.

(c) Modulation of the r.f. by the sub-carrier (ph.m.)

The modulation index (a known characteristic magnitude), given by

$$\begin{aligned} m &= \Phi_m \text{ (the maximum phase shift)} \\ &= \frac{\Delta f \text{ (frequency shift)}}{F \text{ (modulating frequency)}} \end{aligned}$$

is between 0·5 and 0·9 radians.

(d) Binary information rates.

The rate is expressed in digits per second. As a rule, the control system does not require many data to be transmitted, for example, one digit or a few digits per second will suffice. On the other hand, transmission must be extremely reliable, having a digit error factor of, say, 10^{-5}.

The telemetry data are richer in information since these constitute the very purpose of the mission, the rate being a few hundred digits per second, say, at the moment. The acceptable error factor, however, can rise to 10^{-3} or even 10^{-2}.

Obviously, these rates will be improved as techniques improve, and, with rates of the currently feasible order of magnitude, ranges are expected to attain the limit of the solar system. But it is important to know how to use these systems, so as to determine the maximum possible information rate over a given link.

As we have already seen, the information is transmitted in the form of 'digits' (zeros or ones) which we shall assume to be equiprobable for the time being. In information theory, this corresponds to the simplest transmission mode, so that each digit corresponds to a unit of information (or information 'bit' transmitted). Again in accordance with information theory, the quality of the link is characterized by a criterion, known as the Shannon criterion, β, and expressed as

$$\beta = \frac{E}{N_0} \qquad \ldots\ldots(7)$$

where E is the energy (in joules) developed on each bit (and in this case too, on each digit) and N_0 is the noise spectrum density (watts of noise per hertz of bandwidth).

With coherent p.s.k., if P_2' is the power (in watts) developed on the sub-carrier transmitting the digits, and if R is the number of digits transmitted per second, then

$$P_2' = E \times R \qquad \ldots\ldots(8)$$

Further, in view of the modulation indices used, P_2' can be taken to represent roughly half the power received on the r.f. carrier (P_2).

As for the noise spectrum density around the sub-carrier, its value is twice that of the spectrum density obtained with r.f. (due to the two sidebands in the phase modulation by the sub-carrier).

We may therefore write:

$$N_0 = 2KT_K \qquad \ldots\ldots(9)$$

The error probability in digits transmitted by a coherent p.s.k.-modulated sub-carrier, in the presence of noise of uniform density N_0, has been computed by various authors.

It may be written:

$$P_e = \tfrac{1}{2}\left[1-\operatorname{erf}\left(\frac{E}{N_0}\right)^{\frac{1}{2}}\right] = \tfrac{1}{2}[1-\operatorname{erf}(\beta)^{\frac{1}{2}}] \qquad \ldots\ldots(10)$$

where the 'erf' function is the Gaussian error function. P_e can then be plotted against β. The resulting curve is shown in Fig. 2 in which β is expressed in decibels.

For a given link and digit transmission rate R, such a curve will give the digit error probability or, for a given error rate, the transmission performance level to be achieved. In fact the curve gives a value of

$$\beta = \frac{E}{N_0} = \frac{P_2'}{RN_0}$$

making it possible to operate alike on the performance of the r.f. link (which will be reflected in the (P_2'/N_0) ratio) and on the information transmission rate (the number of digits or bits per second). In the case of a poor signal/noise link, it is possible, in other words, to determine the extent to which the number of digits transmitted per second must be reduced accordingly.

With an error factor of 10^{-3} over a Mars-to-Earth link, for instance, it would currently be possible to transmit at the rate of some one hundred digits per second. The Earth-to-Mars command system requires

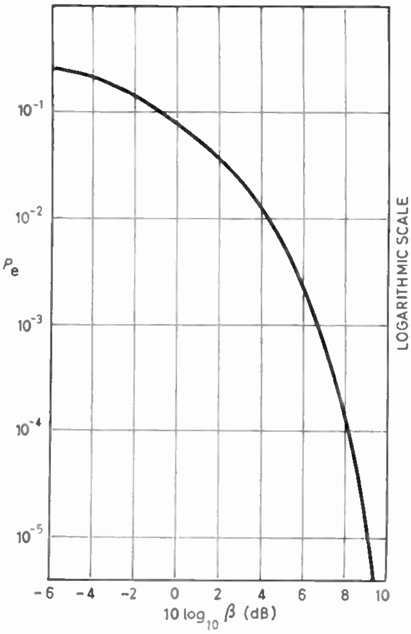


Fig. 2. Error probability in coherent p.s.k. system.
$\beta = \frac{E}{N_0}$ and P_e is the error probability per digit
Note: In p.c.m. 1 digit = 1 bit of information

an appreciably lower error factor (10^{-5}), so that it would be possible to transmit two digits per second with such a link. Table 1 gives the history and future of these information transmission links, showing on-board and ground performance in relation to past, present and future spacecraft missions. Past and anticipated progress in information transmission rates over such links is clearly evident. The rate has risen from 0·00025 bits per second in 1959 to the current figure of 34 bits per second. For the *Voyager* project a rate of 12 000 bits per second is expected. A comparison between these figures shows the accelerating tempo in the advances anticipated for scientific space exploration links.

But, even with the estimated figures for 1973, you will notice how far we are from hoping to use direct television for exploring the Universe. It will still be impossible to televise a planet such as Jupiter and it is a pity since we shall note it is a specially interesting planet.

This is not the place to describe the many scientific results already obtained by the fleet of satellites

Table 1

Performance data for Earth–spacecraft transmissions

	Frequencies used (MHz)	SPACECRAFT Weight (kg)	Power transmitted	Antenna gain and diameter	GROUND STATION Antenna gain and diameter	Noise temperature	Rate R (bit/s)
1959 (*Pioneer IV*)	960	6·5	0·27 W	3 dB	45 dB ($D = 27$ m)	1450° K	0·00025
1962 (*Mariner II*)	960	200	3 W	18 dB ($D = 1{\cdot}3$ m)	46 dB ($D = 27$ m)	250° K	0·7
1965 (*Mariner IV*)	2290	460	10 W	24 dB ($D = 1$ m)	53 dB ($D = 27$ m)	55° K	34
1973 (*Voyager*)	2290	3200	50 W	32 dB ($D = 2{\cdot}2$ m)	61 dB ($D = 65$ m)	25° K	12 000

presently aloft (about six hundred), concerning our ‘environment’, our geodesics and our meteorology; except perhaps to mention that the *ESRO II* satellite, or *Iris*, launched in May 1968, weighs 80 kg and its mission is to study solar astronomy and cosmic radiation. Figure 3 shows the French satellite *DIC* and Fig. 4 shows an interferometer installation.

It should suffice to remember the startling results already obtained on our natural satellite, thanks to the equipment landed on it: the Moon’s physical geography is beginning to become familiar, its relative proximity having made it possible to transmit information bits relevant to exploration missions that can be deferred, thanks to the stability of the lunar

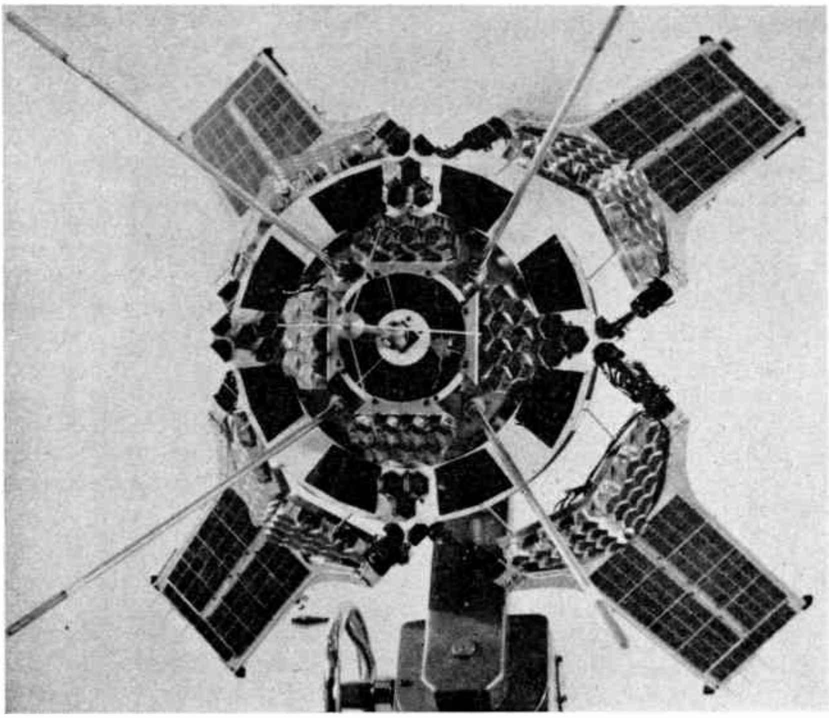


Fig. 3. French satellite *DIC*.

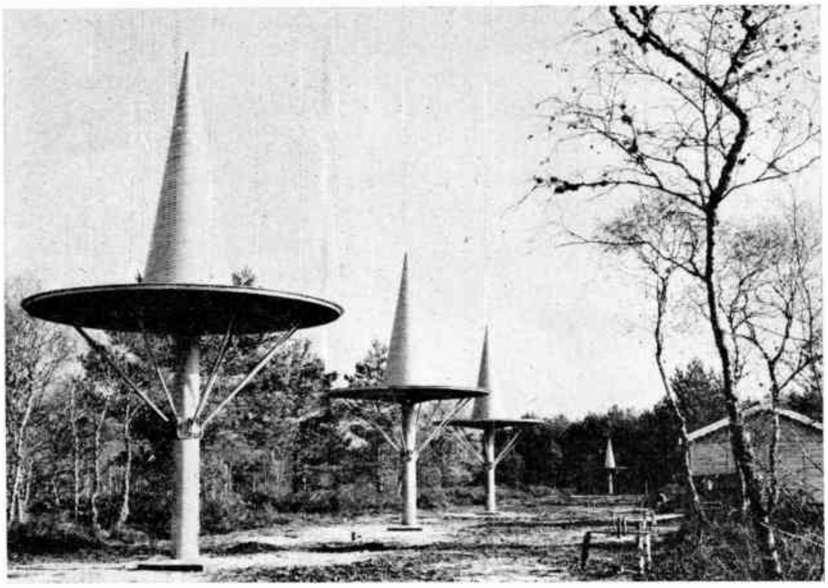


Fig. 4. Interferometer installation.

C.S.F. Photograph

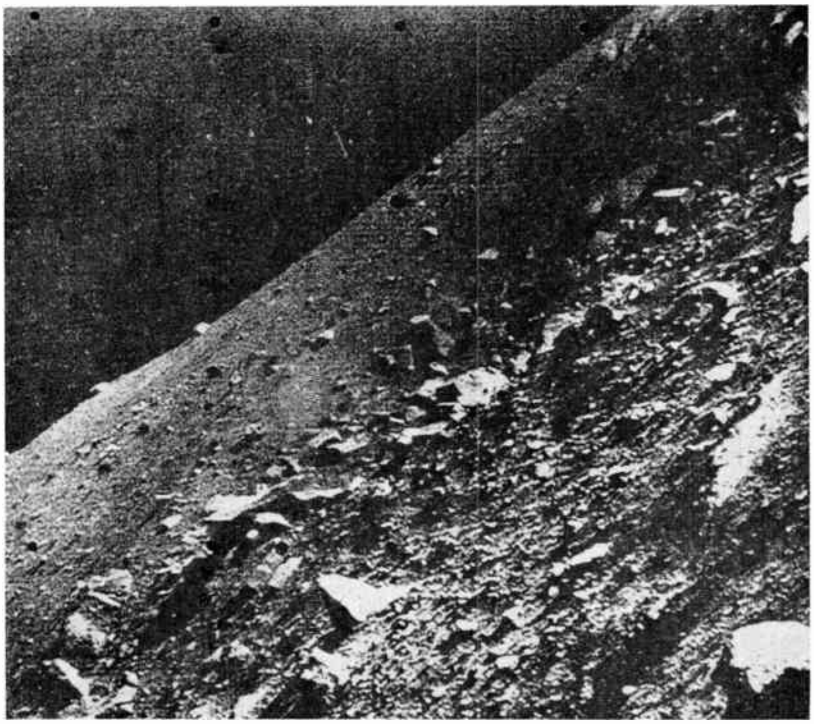


Fig. 5. Lunar soil as photographed by N.A.S.A. from *Surveyor III.*

installations. The lunar soil, with its pebbles, its crust, its dips—shown in Fig. 5—is now fairly well known and can even be simulated in terrestrial laboratories. The Moon's soil even reveals grooves resembling former water courses. The poet Maxwell might well have felt depressed to note that, on the basis of his own theories, that romantic star had been reduced to the state of a lump off the planet Earth that had lost its means to breathe. True, we too will follow suit ultimately, though for different reasons.

It is also interesting to note an observation that contradicts what we were taught a little too hastily: it was thought that in the stellar void the sky would appear black with stars visible in full daylight as brilliant points of light against a velvet background. This is not so, and the sky is very much the same as it appears from the Earth, which must have been a bit upsetting for the many space travel enthusiasts who like to navigate with the naked eye.

Finally, so-called stationary satellites, orbiting 36 000 kilometres above the Earth, can also be used for such purposes as observing our atmosphere for weather forecasting: in fact it is remarkable to note that photographs taken from such a distance with 105° wide-angle lenses provide details smaller than the system's theoretical resolving power.

2.3. *Communication Satellites*

This transition suggests we should take stock of the present situation regarding terrestrial communication satellites.

The trend is towards the unique solution of 'geostationary' satellites, since the drawback arising from the lag in transmission due to the propagation time (0·45 second) can be corrected by an echo suppressor at the cost of greater complication in the ground installations. These satellites enable about one-third of the Earth's surface (an angle of 17°) to be 'sprayed' permanently, and almost the same area always. However, the launching and stabilization problems involved are particularly arduous because the spacecraft must be spin-stabilized with its axis always pointing at the Earth. This problem, however, is considerably facilitated by the use of geo-stationary satellites.

The weight of the satellites alone can be divided into several categories: 110–150 kg (with a maximum of up to 250 kg); 350–500 kg; 1300–2000 kg.

At present, the on-board equipment is powered by solar cells which can provide up to a few tens of watts. On the other hand there is still some uncertainty about their useful life in view of the radiation and particle impacts to which they are subjected.

Looking further into the future, nuclear energy will have to be used when higher powers are needed (e.g. the SNAP project). Some programmes are described in Mr. Bedford's paper already mentioned.

The radio frequencies used at present range from 2000 MHz to 6000 MHz, these bands being divided between space and terrestrial microwave links, which are not normally troublesome since interference is avoided by the antenna directivity.

Here are some examples of the frequency bands used:

Intelsat	Up	6000 MHz
	Down	4000 MHz
Symphonie	Up	6925/6425 MHz
	Down	2700/4200 MHz

The total frequency interval provided makes possible the use of multiple carriers which are themselves modulated with a spectrum coverage of several tens of megahertz.

In the equation given before for the energy balance-sheet of the link (eqn. (3)):

G_1 the antenna gain on the satellite, will be around 15 dB which explains the need for stabilization in order to maintain correct orientation;

G_2 the gain on the ground antenna, could be say 40 dB;

L_S the propagation attenuation is around 190 dB in the case of a stationary satellite 36 000 km aloft, and with the above range of frequencies used.

Thus, for 10 watts emitted from the satellite, the power picked up on the ground is:

$$P_r = 10+15+40-190 = -125 \text{ dBW}$$

This received power, which is of the order of 10^{-12} watts to 10^{-13} watts is comparable with that received over the most difficult terrestrial links using conventional microwave communications systems.

But because of the large number of telephone communications to be handled and the needs imposed by the multiple-access requirement, the input noise at the ground station receiver end must be reduced as much as possible, which is facilitated in the case of a geo-stationary satellite because the antennae, which are aimed at the sky, pick up thermal background noise which is low in comparison with what they would pick up if they were aimed at a much 'hotter' terrestrial target. This is the reason why devices with low inherent noise, such as parametric amplifiers or masers, are used. The up-to-date stations, which at present tend to be used more and more for tracking spacecraft, employ 53-dB gain Cassegrain antennae with noise temperatures of 10° K when aimed at the zenith: the associated masers have a natural temperature of 10° K, with the combined waveguide and antenna this figure reaches 55° K.

The foregoing refers to the satellite-to-Earth trip, called the 'down' trip.

The 'up' trip (Earth-to-satellite) is a less difficult problem, since much higher powers and directivities can obviously be used at the ground transmitter. This means that the energy balance-sheet is much more favourable, but on the other hand a satellite-borne repeater must be capable of handling traffic between several transmitting stations and several receiving stations, for very heavy traffic over a single link would seldom arise, which explains the term 'multiple access'.

This means that several transmissions originating from different points on the planet can be accommodated by the satellite's receiver and sent back to Earth, thus making it necessary to use satellite relaying equipment with the ability to accommodate 'multiple carriers' without causing intermodulation between the carriers. Naturally, each of these carriers is modulated in turn, but by a single process utilizing multiple telephone channels with frequency modulation, or, at some subsequent stage, by utilizing coded pulse transmissions for the different telephone channels.

In order to ensure optimum conditions for intermodulation, the levels of the different transmissions must be adjusted with the utmost care, having regard to possible variations in the propagation attenuations, so as to make sure that the transmissions from the different points on the globe do in fact reach the input end of the satellite's repeater at the required level.

Now, where do we stand in practice?

It is common knowledge that all the telecommunications satellites so far orbited, including those which are no longer in existence, have been either American or Russian, and the yardstick of progress has been the number of channels available and the nature of the ground installations, whose complexity inevitably increase with the number of satellites aloft. Thus, since *Early Bird* (*Intelsat I* (1965), 46 kg) with its 240 channels and an initial cost of \$3400 to \$5400 per hour, and *Blue Bird* (*Intelsat II* (1967), 103 kg), it is interesting to take quick stock of the situation in Western Europe, which is constantly evolving in terms of political factors. Satellites can be used for worldwide or regional communications, the term 'region' being taken in its broadest sense, for instance the whole of Western Europe, i.e. beyond the limits of the Common Market and England.

In view of the technical and technological advance of the United States, does it make economic sense to have a European system at a time when Intelsat exists, along with the Interim Communications Satellite Committee and its manager, COMSAT? At present 62 governments and telecommunication entities are members of Intelsat; 16 ground stations exist in 11 different countries; by the end of 1968 three new stations will become operational in Latin America, and by the end of 1969 there will be 50 stations operating, and over 70 by the end of 1975. Each *Intelsat III* satellite (end of 1968) will have about 1200 two-way telephone channels with multiple access. Subsequently, each *Intelsat IV* will offer 5000 to 6000 two-way telephone channels, thus making the many different facets of modern communications systems (telephony, television, inter-computer links, aeronautical communications) possible within a few years. Following the enormous sums spent on research in the U.S. (\$5000 M in 1967 for NASA alone), it can be estimated that each ground station will cost \$3·5 M, and an *Intelsat III* satellite \$32 M, with the annual cost per unit channel per satellite decreasing in the proportion of ten to one between 1964 and 1976.

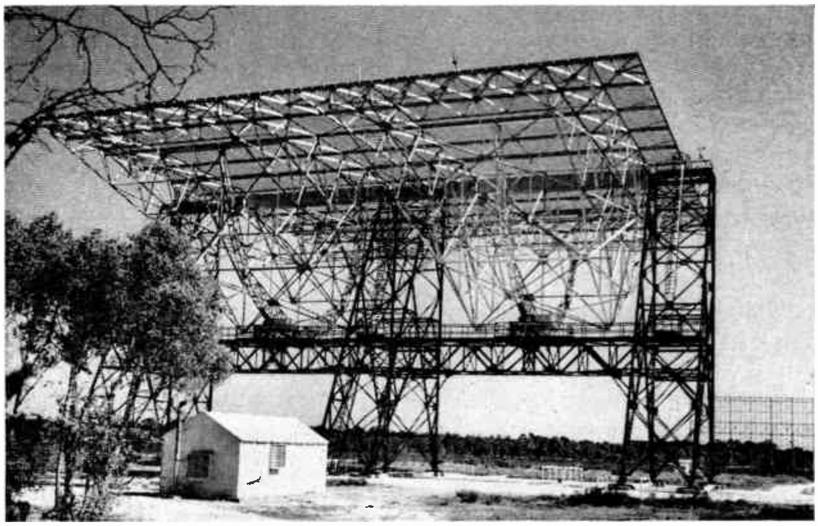
Economically speaking, therefore, the prospect of inserting itself into the global Intelsat system might seem attractive to Europe. There are two things to consider, however: first, space technology enables a national industry to make significant strides forward, well beyond the context of space applications. It is clear that the risk which lies in the Intelsat solution is that those European industries which are allowed to participate—to an extent still to be determined but which is hardly encouraging in view of the contracts obtained so far—will not be required to produce complete systems but only portions that are not likely to involve the most up-to-date techniques. This risk could perhaps be reduced if the United States were to realize that it is in their interests to help Europe develop the latest technologies, as it might otherwise withdraw into itself or turn to the East for other solutions. The second thing to be considered, which holds true whatever the solution adopted, is that Europe cannot afford to be exclusively dependent on other nations in so vital a sphere as telecommunications if it wishes to keep control of the factors that will determine its economic expansion.

The question therefore arises as to whether or not, in view of the considerable delay already incurred, Europe can stand the cost of a regional programme which would be limited by the means at its disposal. Reports on the subject have already been made, notably that of the European Space Conference's Advisory Committee on Programmes. According to this report, an annual sum of \$125 M would be needed to cover the cost of launchers, research, development and construction of such a European system. This represents about 1% of the total annual expenditure earmarked in Europe for communications utilizing electronic systems. The 'regional' system proposed consequently seems reasonable; on the other hand there should be no further delay, for as it is, the system could not become operational before 1972–75, by which time the 'state of the art' will have evolved. True, we must mould a Europe, if only a technological one, but that is another story.

3. Natural Radio Emission from Space

Maxwell's theories apply to all forms of electromagnetic radiation and hence to emissions from celestial bodies. Yet astronomers observed the stars only with the help of the light radiation which their instruments could process, and of course this brought fruitful results.

But in 1889 Hertz discovered 'radioelectric waves' and, on the basis of Maxwell's equations, proved that



C.S.F. Photograph

Fig. 6. Nançay antenna.

they were of an electromagnetic nature, i.e. that there was no fundamental differences between 'radio' waves and the visible waves. Almost immediately afterwards, Sir Oliver Lodge in Britain and Deslandres in France asked themselves whether the Sun emitted this new form of radiation. As often happens, the technical and technological means then available were not up to answering these intuitions and the few experiments that were attempted between 1895 and 1901 failed.

It was therefore only after receiving facilities—antennae and amplifiers in particular—had been developed that the first observations of natural radio emissions were made, and it is interesting to note that, as happened in the early stages of terrestrial radio communications, engineers and 'amateurs' played an essential role. It was the American engineer Jansky who, at a meeting of the American Section of the International Scientific Radio Union in April 1933, announced the results of his observations on emissions at around 20 MHz from the centre of our Galaxy, and it was left to an American amateur astronomer and radio 'ham', Grote Reber, to build the first true radiotelescope. But it was a radio engineer working with the British Army, J. S. Hey, who was the first to make systematic observations of metric wave solar emissions, and this was all the more meritorious since his observations were made in 1942.

Since then radio observatory aerial installations have been built systematically either with conventional matched or with interferometers, the first facility of major importance was erected at Jodrell Bank in this country in 1957. To date there are about a dozen installations in all, including the one at Jodrell Bank, plus others at Parkes (Australia), Columbus (U.S.A.), Dulkovo (U.S.S.R.), Nançay (France) (see Fig. 6) and Bologna (Italy). The most recently-proposed antenna will be in Germany, near Altenahr (Eifel), belonging to the Max Planck Institute. The diameter of the mirror will be 100 metres and this new radio observatory will be ready in 1970.

The host of results already obtained have shown that emissions from the heavens can be divided broadly into the following categories:

Emission from the Sun
Emission from the planets
Emission from our Galaxy
Emission or absorption of interstellar matter
Emission from the radiogalaxies and the quasars

I must admit I feel some qualms about discoursing on this subject here at Cambridge, where researchers and lecturers have contributed so much to the development of this new science. Though I shall not be teaching them anything, no synthesis of the theoretical work based on the Maxwell–Lorentz equations can afford, in 1968, to overlook this domain. And we are already in a position to note that most of these celestial emissions have revealed, on a tremendously vaster scale than any of our terrestrial particle accelerators could ever do, a type of emission which is implicit in the Maxwell–Lorentz equations, namely the 'synchrotron effect'. When electrons are accelerated to speeds close to the speed of light (so-called relativistic electrons) and travel through a magnetic field, they emit radiation, usually within the radio spectrum, that can reach considerable intensity. The intensity of the radiation emitted depends on the energy of the electrons, on their number, and on the magnitude of the magnetic field.

Relativistic electrons with an energy distribution of the form

$$N(E)\,\mathrm{d}E = kE^{-\gamma}\,\mathrm{d}E$$

(where $N(E)\,\mathrm{d}E$ is the number of the electrons with energies between E and $E+\mathrm{d}E$ and γ is close to 2·5 near the Earth) emit radiations of a frequency distribution which in this case is

$$S(f) = S_0 f^{-\alpha} \qquad \ldots\ldots(11)$$

where

$$\alpha = \frac{\gamma-1}{2} \text{ (in this case 0·75)}$$

Lastly, and most important for understanding the nature of an inaccessible emission, synchrotron emission is by its very nature polarized.

3.1. *Emissions from the Sun*

The Sun is certainly the star whose structure has been the subject of most radio observations up to now. Radio waves are obviously not emitted under the same conditions as light waves, the only observable variety up to 1930, though it was light waves which enabled the solar mass to be separated into the photosphere, the chromosphere and the corona. The source of radio emission is the plasmas which exist in the Sun and which can be absorbed in turn providing their frequency is less than a certain limit expressed as

$$f = 9\sqrt{N_e}$$

where N_e is the number of electrons per cubic metre in the plasma layer traversed. Thus, depending on the observation frequency chosen, one can reach any given zone in the Sun, so that localization of the emissions, and the latter's intensity, polarization and variation with time represent so much valuable new data. The Sun, whether calm or active, is the source of all the varieties of emission: thermal, plasma oscillations, or oscillation by the synchrotron effect, the latter being accompanied by cosmic rays whose intensity increases with the degree of agitation of the Sun.

3.2. *Emissions from the Planets*

By virtue of their spectrum and intensity, radio emissions from the planets, including the Moon, had been regarded as 'thermal' until 1954, when Burke and Franklin detected emission from Jupiter on 22 MHz: this takes the form of sequences of brief spurts similar to those noted in solar radio noise storms, only more intense. At the other end of the spectrum, other polarized emissions can be observed on decimetric waves. The existence of all these emissions might be explained by the presence of a magnetic field round Jupiter caused by a very widespread ionosphere, the size of the decimetric-wave source being three to four times as great as that of the visible disk.

3.3. *Emissions from the Galaxy*

The emission from our Galaxy was among the first to be observed, though it turns out to be weak because our Galaxy is 'old' in terms of what can now be deduced from observation of the other galaxies. The radio emissions include a continuous synchrotron emission, from which the magnetic field in the galactic disk can be estimated at 10^{-5} to 10^{-6} gauss, extending to about 200 parsecs on either side of the disk's plane (1 parsec = 3·1 light-years).

3.4. *Interstellar Space*

Neutral hydrogen is the major constituent of interstellar gas and produces an emission (or absorption) line on 1420 MHz (21·2 cm) due to the transition between the two possible spins of an electron in the fundamental state of the atom. This emission has been explored right to the very heart of the Galaxy and has enabled the spiral form to be established. The mass of hydrogen has been estimated at $1{\cdot}5\times10^9$ solar masses, but at only 1% of the total mass of the Galaxy. This should make us feel humble indeed, lost as we are on our tiny solar system! (see Fig. 7).

Studies of hydrogen emission from the galaxies and the far reaches of interstellar space are continuing, notably at Green Bank and Nançay.

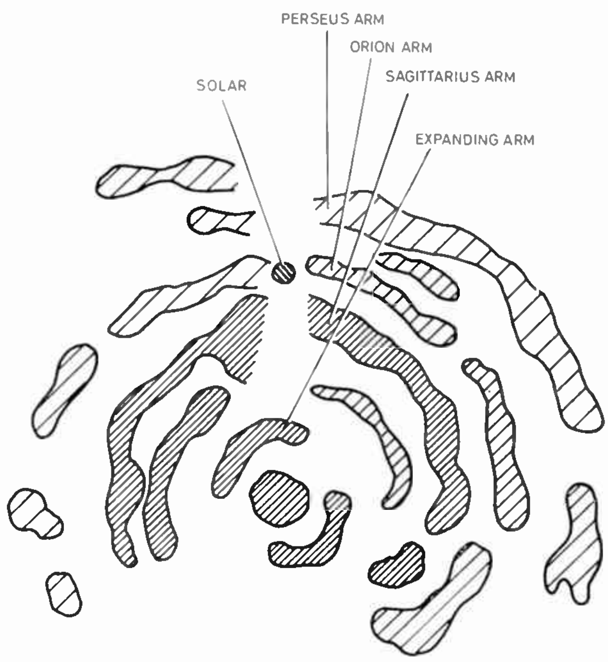


Fig. 7. Structure of our Galaxy.

It should be noted also that absorption lines of the radical –OH have been observed between 1610 and 1720 MHz, another fundamental result from the study of interstellar matter.

3.5. *Radiogalaxies and Quasars*

Radio astronomers have discovered numerous discrete sources of radio emissions from beyond our Galaxy. Cambridge University has played a leading role in locating and numbering them, for Cambridge's third catalogue on 169 MHz, which included 471 sources when it was issued, was the origin for the designation 3 CX. These 'radiogalaxies', whose positions generally coincide with visible bodies at the very limit of optical telescopes, are powerful emitters, no doubt by a synchrotron effect, the radio energy of which attains an important proportion of the visible energy, which does not apply in our case. As a rule, these radiogalaxies have a complex structure with separate nuclei.

It was in 1962–63 that the first observations and measurements were made of the special properties of a very intense radio source (3 C 273), which it had not been possible to attribute to a known galaxy. An Englishman, Hazard, and two Australians, Machez and Shimmins, working at Parkes in Australia, used the occultation of the source by the Moon to establish its structure and precise position. The source had a stellar aspect, was luminous, and its spectrum of optical lines revealed a heavy shift towards the red. About one hundred of these bodies are now known and are called quasi-stellar radio sources, or quasars. They are now the subject of intensive study and have revealed many contradictions: yet to understand the Universe about us it is essential that we should find out everything we can about them. There is reason to believe that the origin of the quasars can be ascribed to explosions of galactic nuclei, and that in time they will be transformed into radiogalaxies and galaxies. But if the shift in their optical spectrum lines, which is invariably towards the red, is due to the Doppler effect, i.e.

$$z = \frac{1 + \dfrac{v}{c}\cos\alpha}{\sqrt{1 - \dfrac{v^2}{c^2}}} - 1$$

then both the rate at which they are receding and their distance are enormous being respectively $0{\cdot}35\,c$ and 500×10^6 parsecs in the case of 3 C 273 (or 1500 million light years.) In view of the relatively rapid variations in their brightness, the sources at such distances must be of comparatively small size, and the energy emitted by these volumes must be enormous, being of the order of 10^{60} erg/s, equivalent to the simultaneous explosion of a million suns, 10% of the energy being radio energy. What the source of such energy could be is not yet understood. The shift in the spectrum lines towards the red might be due to some other cause, in which case the quasars would be closer to us. Figure 8 shows some quantitative and statistical data relative to these sources.

Thus, the observation of radiation on all frequencies from the Universe, the theoretical interpretation of which has hitherto been based on the Maxwell–Lorentz theories, is proving most fertile and is one of the essential elements in broadening man's knowledge.

4. Conclusions

In conclusion, perhaps it is worth noting that in 1868 the elaboration of the mathematical theory of fields consecrated the distinction between Science and

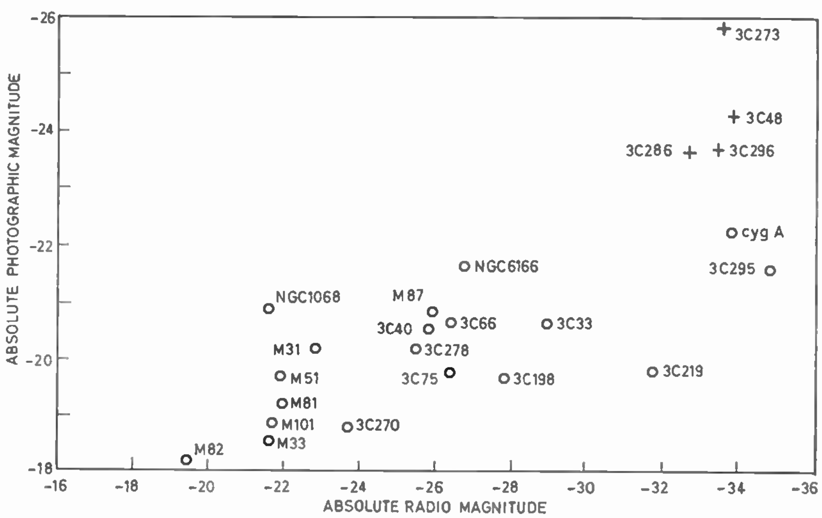


Fig. 8. Relation between radio and optical magnitudes of several sources.

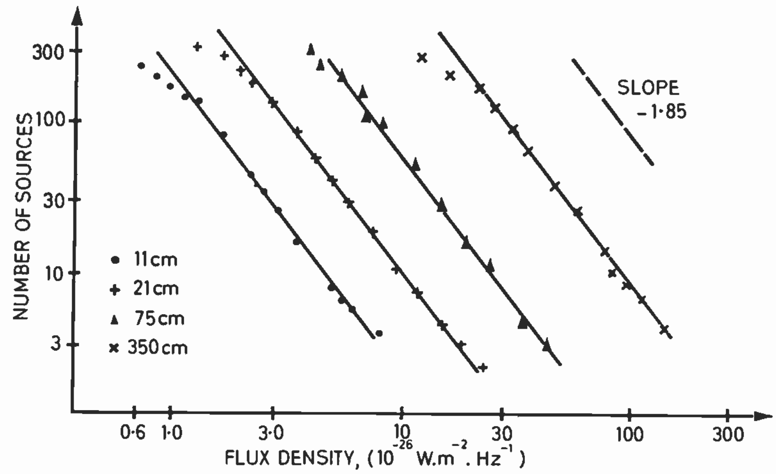


Fig. 9. Number of radio sources vs. flux.

'Natural Philosophy', the latter having still guided the work of Oersted and Faraday, at least initially. So perhaps Clerk Maxwell, whose younger years had been spent in a philosophical environment which was still steeped, in Britain, in the ideas of Kant, would have been sorry to have dealt such a severe blow to what might be termed sentimental scientific philosophy. Following the work of great pure scientists like Ampère, Maxwell and Sir J. J. Thomson, Science ceased, from the end of the nineteenth century onwards, to launch into philosophical considerations like those on the origin of the Universe.

Yet, in spite of ourselves, we find the study of stellar radio sources bringing us face-to-face with problems of this order once more. By virtue of its great sensitivity, radio astronomy is probably alone at the present time in being able to provide information about objects distant enough to have a bearing on the origin of the Universe. Thanks to radio astronomy it is now possible to reach these freshly-exploded 'objects' which are receding in an expanding universe at fantastic speeds equal to half the speed of light in some cases (another quasar, 3 C 147, is moving at $0{\cdot}55\,c$ and is $1{\cdot}6\times10^9$ parsecs away.)

But if the Universe is stationary through a process of continuous creation, the total number of sources in a given volume must remain constant in near and distant regions alike, and this can be observed by studying the number of sources as a function of their intensity. The power received on Earth from a radio source with a spectral intensity $I(f)\mathrm{d}f$ located at distance r is, in frequency and surface units:

$$F(f) = \frac{I(f)}{4\pi r^2}$$

Superior fluxes come from nearer sources inside a sphere of radius r. If the Universe follows a process of continuous creation, the volumetric density δ of the sources is uniform and, therefore, the number of sources N responsible of a flux superior to $F(f)$ will then be

$$N = \frac{4}{3}\pi r^3 \delta$$

For a given frequency, this number will therefore be of the form:

$$N = KF_f^{-1\cdot5}$$

Now the British astronomer Sir Martin Ryle, at Cambridge, found that the curve log N vs. log F for the total of known radio sources is a straight line of slope 1·85 as shown in Fig. 9, and this is confirmed by his Australian colleagues; the slope of the line related to quasars is still greater (2·15) but its interpretation is dubious. It seems, in any case, that weak radio sources are more numerous than theory would indicate: hence there had apparently been a greater number of radiogalaxies in the first days of the Universe . . . some ten thousand million years ago, insofar as such a figure means anything at all, and assuming the Universe to have had its origins in one explosion. It would seem therefore that the premise of continuous creation must be disregarded, which has been contested by Fred Hoyle in this very place, and meanwhile there is still no clear indication of whether the model which governs the Universe is 'parabolic, hyperbolic or elliptic'.

In any event, thanks to Maxwell, we are faced with the riddle again one hundred years later, and that great physicist would certainly have been gratified to note this return to the universality of the mind, in which Cambridge continues to play so brilliant a part.

Manuscript received by the Institution on 22nd July 1968 (Address No. 40).

Of Current Interest . . .

An Institute of Cybernetics

One of the more interesting developments in recent years, fostered it is fair to say by the vast expansion of electronics into fields of computers, automatic control and biological electronics, has been the identification of 'cybernetics' as a discipline. Drawing its theories and techniques from the subjects just mentioned, as well as from communications, the derivation of the word is from the Greek name for a 'steersman'. Brunel, the great engineer, who pioneered canal and railway building in this country, would therefore certainly have welcomed the establishment of an 'Institute of Cybernetics' at Brunel University, Hillingdon, Middlesex.

Dr. Frank George has been appointed first Professor of Cybernetics. Financial support for the Institute has been given by the International Publishing Corporation and further industrial backing is being sought.

Third Annual Report of the Science Research Council

The third annual report of the Science Research Council†, shows that the Council has continued to develop its policies and to establish priorities for a period in which fewer resources are available than are needed to carry out all the programmes which are desirable on scientific grounds. The main lines of policy determined are:

(i) While continuing to devote most of its funds to basic research, the Council will for the next few years still give preferentially increasing support to applied research and will encourage the application in the national interest of the results and by-products of basic research;

(ii) in allocating the severely limited funds which the Government has been able to make available to it the Council will be increasingly selective in supporting particular fields of scientific and technological importance. In expensive fields it will concentrate support in a limited number of laboratories;

(iii) in the present economic situation the Council will restrict the number of academic and professional staff which it supports through research grants with the intention that the trained manpower so released may be attracted into industry or teaching in this country. This is in accordance with the recent Swann Report (the Flow into Employment of Scientists, Engineers and Technologists); and

(iv) the Council will continue to offer substantial support for the postgraduate training of scientists and technologists with increasing emphasis on training relevant to industrial needs.

The Council continues to rely on European collaboration in large programmes such as those of CERN and ESRO.

In the field of space research, subject to the uncertainty over the future ESRO programme, the Council will aim to provide satellite and rocket facilities for selected projects within a budget of about one-fifth of its resources.

In the field of nuclear physics the Council regards participation in the 300 GeV or a similar international project as essential to the proper development of the subject in the late 1970s. It will, therefore, continue to work towards participation in such a scheme while containing its support for nuclear physics within a falling proportion of its total resources.

In 1967–8 the proportion of the Council's funds spent on international organizations (CERN, ESRO and NATO scientific schemes) increased from 23% to 26%, partly as a result of devaluation, while expenditure in SRC laboratories fell from 42% to 39%. The proportion of expenditure incurred in universities remained steady at about 35%.

† Report of the Science Research Council for the year 1967–68. Published for SRC by HMSO. Price 7s. 6d. net.

European Data Communication

Problems raised by the rapid spread of computers linked to telecommunication networks for on-line working were among important questions discussed by the sub-working group of CEPT (Conference of European Telecommunication Administrations) at its meeting in London in July. The meeting, attended by representatives from ten European Telecommunications Administrations, was called to consider the characteristics required for networks for data communications.

Of particular interest and importance was the multi-access computer (especially in the form of computer bureaux) linked to remote keyboard instruments such as teleprinters. A developing need was expected to be that for high-speed data transmission, for example, at the rate of 48 kbit per second recently recommended by the Study Group Special 'A' of the CCITT. Characteristics which a future network should have were discussed in order to be able to cater for a wide range of facilities and types of user.

British Experiments in ESRO I Satellite

Two British Research groups have experiments in the European Space Research Organization's second satellite *ESRO I*—now known as *Aurorae*—in the ESRO–NASA co-operative space research programme. *Aurorae* was launched on 3rd October from the Vandenburg Range in California by a NASA *Scout* rocket. Early reports indicate that the launch was successful and that the satellite has gone into an elliptical polar orbit. This orbit was chosen to give a good distribution of measurements and during the first few months the satellite's sub-orbital track will enable auroral phenomena to be observed throughout the northern winter.

The satellite is stabilized along the magnetic field lines by permanent magnets in the structure.

The SRC Radio and Space Research Station experiments will measure electron flux and energy spectra in the 40–400 keV range, and proton flux and spectra in the 6–30 MeV range.

University College London's Mullard Space Science Laboratory at Holmbury St. Mary, Dorking, which received SRC support totalling over £150,000 for space research during 1968, is making electron temperature and density measurements and studying the composition and temperature of positive ions in the upper atmosphere.

U.D.C. 621.397 : 629.19

Television by Satellite

By

L. H. BEDFORD,

C.B.E., M.A., B.Sc.(Eng.), F.C.G.I., C.Eng., F.I.E.E., Hon.F.I.E.R.E.†

Presented at the Meeting of the Comitato Internazionale di Televisione held in Milan in April 1968.

Summary: The transmitter power requirements for adequate field strength are evaluated for direct broadcasting in the present u.h.f. television bands. It is shown that improved aerials, the use of preamplifiers and wideband frequency modulation will be necessary for practical transmitters in existing or foreseeable space vehicles. The concept of 'indicial bandwidth' is introduced which is defined as the receiver bandwidth for which carrier/noise power-ratio is unity. Satellite considerations of power supply, station keeping and launching costs are discussed. The conclusion is reached that satellite broadcasting can most satisfactorily operate to stations feeding conventional wire broadcasting networks.

List of Symbols

A_e	reception area, m^2
A_t	effective area of (satellite) transmitting antenna, m^2
A_r	effective area of (ground) receiving antenna, m^2
α	angle of elevation to satellite, deg
Ω	solid angle subtended by reception area at satellite, steradians
λ	wavelength of transmission, m
f	carrier frequency of transmission, Hz
T	system noise temperature, degK
k	Boltzmann's constant ($= 1{\cdot}38 \times 10^{-23}$ J/degK)
r	distance from receiver to stationary satellite, m
P_t	satellite transmitted r.f. power, W
P_c	carrier power into receiver, W
B_0	indicial bandwidth ($\equiv P_c/kT$), Hz
B_1	transmitted r.f. bandwidth, Hz
B_2	i.f. bandwidth, Hz
B_3	information bandwidth, Hz
f_d	maximum frequency deviation (for f.m.), Hz
M	modulation index ($\equiv f_d/B_3$)
ρ_t	threshold power-ratio
β	f.m. feedback factor

1. Communications Aspects

The term 'Television by Satellite' is open to many interpretations, of which the most extreme in technical difficulty is the case of the direct television broadcaster, that is to say, a satellite transmitting directly into home receivers. This concept is again open to various shades of meaning of which the ultimate implies the use of existing receivers without adaptation.

Although this case proves to be much beyond 'state of the art' possibilities, it will be convenient to commence with it as a reference point and to work backwards from it to the more immediate possibilities.

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

1.1. *Direct Broadcasting to Existing Receivers*

The stipulation of an 'existing' receiver saves us the necessity, at least in the first instance, of enquiring into such quantities as required signal/noise ratio, receiver noise factor, or frequency assignment. These considerations are jointly written into the F.C.C. definition of a 'grade A service' field-strength which, for the u.h.f. band, say 470–890 MHz, is stated as 5010 μV/m (74 dB above 1 μV/m). We can immediately convert this to a power flux density; the impedance of space being 120π ohms:

$$1\ \text{V/m} \equiv \frac{1}{120\pi}\ \text{W/m}^2$$

$$\text{Thus } 5010\,\mu\text{V/m} \equiv \frac{1}{120\pi} \times (5010 \times 10^{-6})^2\ \text{W/m}^2$$

$$\simeq 6{\cdot}6 \times 10^{-8}\ \text{W/m}^2 \qquad \text{......(1)}$$

If, then, we wish to illuminate a small area $A_e(\text{m}^2)$ of the earth's surface, and if we can arrange that the directivity of the satellite transmitting antenna is such as to concentrate the whole transmitter power on to this area, the required transmitter power W_1 for grade A field-strength is simply

$$6{\cdot}7 \times 10^{-8} A_e\ \text{watts} \qquad \text{......(2)}$$

This reckoning, which of course is for the case of a single television channel, is for normal illumination of the area A_e, that is to say from an overhead satellite. However, since in the present context only geostationary satellites are eligible for consideration, an overhead satellite can occur only for an equatorial receiving site. In any other case, the angle of elevation α to the satellite will be less than 90°, and will in fact have a maximum possible value α_0 approximating to the co-latitude.

Table 1

Region	U.K.	U.K., France, Germany, Italy and Spain	U.S.A.	Australia
(2) Latitude	50° to 58°	36° to 58°	25° to 48°	−11° to −38°
(3) Longitude	10°W to 2°E	10°W to 18°E	125°W to 80°W	113°E to 153°E
(4) Area: square miles	$2{\cdot}32 \times 10^5$	$2{\cdot}12 \times 10^6$	$3{\cdot}94 \times 10^6$	$3{\cdot}01 \times 10^6$
square km	$6{\cdot}0 \times 10^5$	$5{\cdot}34 \times 10^6$	$10{\cdot}2 \times 10^6$	$7{\cdot}8 \times 10^6$
Earth's surface	$1{\cdot}18 \times 10^{-3}$	$1{\cdot}05 \times 10^{-2}$	$2{\cdot}0 \times 10^{-2}$	$1{\cdot}53 \times 10^{-2}$
(5) Mean angle of elevation, α_0	28·0°	37·2°	48·7°	62·1°
(6) Solid angle subtended at satellite, Ω (steradians)	$1{\cdot}92 \times 10^{-4}$	$2{\cdot}28 \times 10^{-3}$	$5{\cdot}6 \times 10^{-3}$	$5{\cdot}5 \times 10^{-3}$
(7) R.f. power requirement, W_3 (kW)	37·7	432	1027	924
(8) Minimum r.f. power requirement for $\alpha = 10°$, W_{10} (kW)	13·9	124	237	182
(9) E.r.p.: megawatts	1234	1186	1150	1054
dBW	90·9	90·7	90·6	90·2
(10) Satellite antenna gain (dB)	45·0	34·4	30·5	30·5
(11) A_t/λ^2	2600	220	89	90
(12) $\sqrt{A_t}$ (m); for $f = 890$ MHz	17·3	5·0	3·2	2·3

At first sight it might be thought that this obliquity of illumination would result in a loss. In fact, as explicitly pointed out by Hilton,[1] the contrary is the case. This is because, from the satellite point of view, the area to be illuminated is not A_e but the projection of this area normal to the line of sight, namely $A_e \sin\alpha$.

We accordingly amend eqn. (2) to

$$W_2 = 6{\cdot}7 \times 10^{-8} A_e \sin\alpha \quad \text{watts} \ldots\ldots(3)$$

where α is the elevation of the satellite for the centre of the area A_e concerned. Equation (3) notes the strange result that it does not pay to station the satellite on the longitude of the reception area but as far away from it as possible! However, the following qualifications accompany this statement:

(a) The basic implication is that it must be possible to follow up reduction of α with a corresponding reduction of antenna beam width in the direction concerned.

(b) Receiving antennas will have their collecting areas normal to the line of sight.

(c) Atmospheric and terrain losses set a lower limit to the minimum value of α. In New York City type of terrain, α must be as high as possible, but for moderately open terrain α may perhaps be as low as 10°.

Equations (2) and (3) are based on the somewhat naïve assumption that it is possible to distribute the whole transmitter power uniformly over a given reception area. This would imply, for an area subtending solid angle Ω (steradians) at the satellite, an antenna gain of $4\pi/\Omega$ for the whole of this solid angle and zero everywhere else. In practice this is not possible. The practical case is covered to a fair degree of approximation by stating that if Ω is the solid angle subtended by the −3 dB contour, the axial gain of the antenna is approximately $2\pi/\Omega$. (The practical losses thus amount to more than 3 dB.)

Accordingly, we must amend eqn. (3) to

$$W_3 = 13{\cdot}4 \times 10^{-8} A_e \sin\alpha \quad \text{watts} \ldots\ldots(4)$$

It is interesting to note that these equations do not explicitly involve two quantities which we might reasonably have expected to see, namely the frequency and distance of the satellite transmitter. In fact the latter quantity occurs implicitly as defining the antenna directivity required, and the frequency occurs implicitly in the F.C.C. field-strength requirement as expressing the receiver 'state of the art'.

We may now use eqn. (4) to evaluate the transmitter power required for a number of cases. Table 1 shows example regions listed in columns (2) to (5). In rows (2) and (3) we specify by means of latitude and longitude lines appropriate spherical quadrilaterals circumscribing these regions, and we will refer to these quadrilaterals as the receiving areas.

Row (4) gives the areas of these quadrilaterals in square miles, square kilometres and proportion of earth surface.

Row (5) gives the angle of elevation α_0 from the centre of area to the satellite, assumed stationed on its mean longitude.

Row (6) gives the solid angle (steradians) subtended by the reception area at the satellite.

Row (7) gives the power requirement on the basis of eqn. (4), again assuming the satellite to be stationed on the mean longitude of the reception area.

Row (8) gives the minimum power assuming the satellite to be off-stationed in longitude so as to give angle of elevation of 10° to the satellite.

Note that these powers are 'actual', *not* 'effective radiated', that is to say *not* 'e.r.p.s'. To avoid possible confusion the latter are shown in row (9). These e.r.p.s are shown in megawatts (!) and in dBW. It is somewhat revealing to note that the e.r.p.s vary by only 1 dB over the whole range of examples, the variation being in fact solely that of the 'elevation factor'.

Row (10) shows the required satellite antenna gain.

Row (11) shows the area of the satellite antenna in 'square wavelengths'.

Row (12) shows the square root of the antenna area for a wavelength of 0·34 m (f = 890 MHz).

These figures for r.f. power and antenna size are quite appalling. If not actually impossible, they are at least sufficiently far from the practical state of the art as to render further evaluation here unprofitable. The primary use of these figures is as a basis of reference. In a secondary role, they serve to show how far from reality is the popular interpretation of direct television broadcasting by satellite.

1.2. *Improvements in Receiving Antennas and Noise Temperatures*

To seek a reduction of the required power we must retreat from the concept of the use of existing receivers and consider first of all improved receiving antennas and noise temperatures. As a preface to this, we must first interpret the F.C.C. requirement in terms of these quantities.

Denoting by A_r the collecting area of the receiving antenna and assuming that Grade A service implies a 40 dB peak-signal/noise ratio at video frequency, or 44 dB at r.f.:†

Peak r.f. signal power into receiver

$$= 6{\cdot}7\times10^{-8}A_r \text{ watts}$$

Noise power at receiver input = kTB.

Hence

$$\frac{6{\cdot}7\times10^{-8}A_r}{kTB} = 2{\cdot}5\times10^4$$

$$\frac{A_r}{T} = 2{\cdot}5\times10^4 kB\times\frac{1}{6{\cdot}7}\times10^8$$

$$= 2{\cdot}5\times10^4\times1{\cdot}38\times10^{-23}\times6\times10^6\times\frac{1}{6{\cdot}7}\times10^8$$

$$= 3{\cdot}1\times10^{-5}$$

For the temperature T we may assume 7000°K, whence

$$A_r = 0{\cdot}216 \text{ m}^2$$

These figures for T and A_r we take to define the 'reference case'. To see what this means in terms of receiving antenna gain, we note that

$$G_r = \frac{4\pi A_r}{\lambda^2}$$

$$G_r(\lambda = 0{\cdot}34 \text{ m}) = \frac{4\pi\times0{\cdot}216}{(0{\cdot}34)^2} = 23{\cdot}5, \text{ say } 14 \text{ dB}$$

Also

$$\frac{G_r}{T} \simeq 3{\cdot}36\times10^{-3} \quad\text{and}\quad 10\log\frac{G_r}{T} \simeq -25 \text{ dB}$$

These 'G/T' figures are for ease of reference; but note that A_r/T is the operative quantity in the present application, not G_r/T.

To explore the possibilities of radical improvement we must now analyse the situation rather more fundamentally.

The power P_r into the receiver due to power P_t at the transmitter is clearly

$$\frac{P_t}{4\pi r^2}G_t A_r$$

while the noise power is kTB.

Hence

$$\text{carrier/noise power ratio} = \frac{G_t}{4\pi r^2}A_r\frac{P_t}{kTB} \quad\text{......(5)}$$

In this equation, Boltzmann's constant k and the distance to the stationary satellite, r, are constants of Nature. G_t is determined by the geometry of the receiving area, so the only parameters available for manipulation are A_r, T, B and P_t; the object of manipulation is to minimize the latter.

Considering these quantities in turn, we start with A_r. It is thought that by elaborating the Yagi array within the limits of a practical installation it may be possible to increase A_r from the reference case value by a factor of 3 dB. This is clearly inadequate to transform the situation, so we look next at the alternative of a parabolic reflector or 'dish' antenna. We will take as the limit for domestic practicality an

† This discounts any degradation by the detection process, which according to some writers may amount to some 4 dB.

effective diameter of 1 m; much larger dishes have indeed been proposed, but these are considered non-domestic and appropriate rather to community reception or the like. The 1 m dish gives

$$A_r = \frac{\pi}{4}\ \mathrm{m}^2$$

an improvement of approximately 6 dB over the reference Yagi array.

We next look at the system temperature T. Referring to Fig. 1, and still staying at 890 MHz for the transmission frequency, we see that the substantial improvements shown in Table 2 are possible.

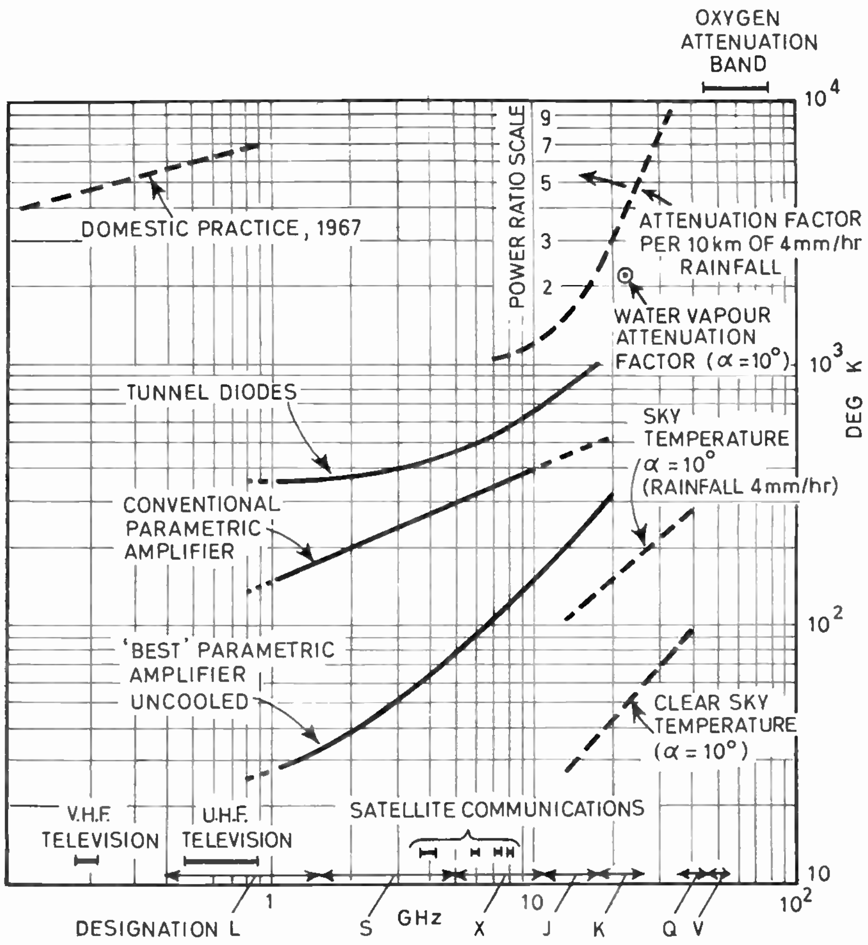


Fig. 1. Frequency band allocation, 'state of the art', and natural limitations based on Refs. 2, 3, 4.

Table 2

Arrangement	System temperature (deg K)	Antenna collecting area, A_r, (m²)	E.R.P. requirement	R.F. power requirement for U.K. area
Reference case	7000	0·216	1200 MW	37·7 kW
Tunnel diode head amplifier	350 350	0·216 0·786	60 MW 16·5 MW	1·85 kW 508 W
Conventional parametric amplifier	145	0·786	6·86 MW	210 W
'Best' parametric (uncooled) amplifier	27	0·786	1·27 MW	38 W

While these later figures for required transmitter power begin to look somewhat inviting, it must be said that neither of the two parametric amplifier solutions can be regarded as 'domestic', so that the best we have reached is 508 watts per video channel for U.K. coverage and proportionally more for the other cases. To reach this situation the investment at the home receiver is a 1 m diameter dish and a tunnel diode preamplifier; by no means inconsiderable items.

It should be re-emphasized that, as shown by eqn. (5), frequency does not enter in a fundamental manner into the reckoning but only in a 'state of the art' fashion. Figure 1 depicts the situation and shows that increase of transmitter frequency deteriorates the situation at the receiver end; this is because for purely 'state of the art' reasons, system temperature increases with frequency. Furthermore, change from the u.h.f. band involves another receiver penalty, namely, a new or additional frequency changer stage. The benefit of increased frequency is found entirely at the transmitting (satellite) end where it offers a much needed reduction of antenna size.

Fortunately, however, major system gains are still possible once the necessity of an adaptor at the receiver end is accepted. So far, in the interests of keeping as close as possible to a standard domestic receiver (though not very close at that!), we have stayed with standard amplitude modulation. But this is not a preferred form of modulation for satellite working where, because of the great expense of satellite power, one is willing to trade frequency bandwidth for power economy. Information theory indicates the possibility of this, and the best known and most easily 'domesticized' embodiment of the principle is frequency modulation.

1.3. *Wideband Frequency Modulation*

We now proceed to an analysis of the system using wideband frequency modulation.

As a preliminary it will be useful to refer to a notational point. The quantity which turns out to be fundamental in this and all such analyses is P_c/kT, where P_c is the carrier power into the receiver, T is the receiving system noise temperature (all noise sources being assumed thermal in character), and k is Boltzmann's constant ($= 1{\cdot}38 \times 10^{-23}$ J/deg K). Noting that P_c/kT is in the nature of a system index, and has the dimensions of a bandwidth, the author christened it the 'indicial bandwidth, B_0'. It is of course a system quantity including both transmitter and receiver parameters; and is in fact the receiver bandwidth for which the carrier/noise power-ratio is unity.

It turns out then that there are the following four bandwidths to be considered:

B_0 indicial bandwidth

B_1 transmitted r.f. bandwidth

B_2 i.f. bandwidth

B_3 information or video bandwidth

We also introduce:

f_d the maximum frequency deviation $= MB_3$,

where M is the modulation index.

Then Carson's rule gives:

$$B_1 = 2(B_3 + f_d) = 2B_3(1+M) \qquad \ldots\ldots(6)$$

$$\frac{B_0}{B_2} \equiv \text{working threshold power ratio} = \rho_t \ \ldots\ldots(7)$$

We must not ignore the principle of 'f.m. feedback' as a potential contributor to the 'decibel hunt'. To allow for this we define a feedback factor β such that $\beta = 0$ for zero feedback and $\beta = 1$ for (hypothetical) complete frequency following. We may then regard all frequency deviations as scaled down after feedback by the factor $(1-\beta)$ and

$$B_2 = 2(B_3 + \overline{1-\beta} f_d)$$
$$= 2B_3(1 + \overline{1-\beta} M) \qquad \ldots\ldots(8)$$

Finally, the video signal/noise power-ratio is shown in the Appendix to be:

$$\rho_v = 12 \frac{B_0}{B_2} M^2(1 + \overline{1-\beta} M) \qquad \ldots\ldots(9)$$

Equations (6), (7), (8) and (9) are four simultaneous equations among eight variables of which only three are prescribed. If we assign values to any one of the five remaining variables we obtain a unique solution for the other four. The most convenient variable to select for this 'parametric' treatment is β, for which we shall here consider only the extreme and median cases. These are set out in Table 3.

The bracketed rows in Table 3 relate to the hypothetical case of complete frequency following ($\beta = 1$). These figures, shown in parentheses, are in the nature of hypothetical limits.

In fact with such high bandwidths as are here involved the practicality of any f.m. feedback may be questioned. There is fortunately an alternative, the so-called phase-locked-loop receiver which in some respects corresponds to the case of full frequency following. The phase-locked-loop receiver figures are shown in rows 13 and 14 of Table 3.

A property of the phase-locked-loop receiver, as expounded by Haggai,[11] is that, unlike the normally accepted f.m. receiver, the system works linearly down to a threshold of 6 dB; and even below this threshold the performance degrades smoothly rather than catastrophically. In Table 3 we see that the total system gain by the use of this system is, for example,

$$10 \log \frac{1000}{317} \simeq 5 \text{ dB}$$

which in the present context is highly valuable, if not revolutionary.

Having determined the required indicial bandwidths as in Table 3, we now have to translate these into effective radiated powers P_e for comparison with Table 1. The conversion formula reduces to

$$P_e = B_0 kT \times 16 \left(\frac{r}{d}\right)^2 \qquad \text{......(10)}$$

where r is the distance of the satellite

and d is the diameter of the receiving dish.

For a geostationary satellite and a 1 m diameter dish this becomes

$$P_e = B_0 \times 28{\cdot}8 \times 10^{-8} T \qquad \text{......(11)}$$

Table 4 shows the effective radiated powers for several systems.

1.4. *The Sound Channel*

From the Appendix

$$\left[\frac{\text{r.m.s. frequency deviation due to tone}}{\text{r.m.s. frequency deviation due to noise}}\right]^2$$

$$= 3 \frac{B_0}{B_2} M^2(1+\overline{1-\beta M}) \qquad \text{......(12)}$$

We will assume a tone/noise requirement of 51 dB (55 dB less 4 dB correction for pre-emphasis). Thus we require

$$3 \frac{B_0}{B_2} M^2(1+\overline{1-\beta M}) > 10^{5{\cdot}1} \qquad \text{......(13)}$$

or

$$\rho_t M^2(1+\overline{1-\beta M}) > 4{\cdot}2 \times 10^4 \qquad \text{......(14)}$$

We accordingly obtain the results shown in Table 5.

Comparing Table 5 with Table 3 we see that vastly lower values of B_0 are required for the sound channel,

Table 3 Indicial bandwidths required under various conditions

$10 \log \rho_v$ (dB)	$10 \log \rho_t$ (dB)	β	M	B_1 (MHz)	B_2 (MHz)	B_0 (req) (MHz)
40	13	0	3·17	50·0	50·0	1000
40	13	½	3·8	57·6	34·8	696
[40	13	1	6·46	77·5	12·0	240]
40	10	0	4·05	60·6	60·6	606
40	10	½	4·9	70·8	41·4	414
[40	10	1	9·1	121·3	12·0	120]
43	16	0	3·17	50·0	50·0	2000
43	16	½	3·8	57·6	34·8	1392
[43	16	1	6·46	77·5	12·0	480]
43	13	0	4·05	60·6	60·6	1212
43	13	½	4·9	70·8	41·4	818
[43	13	1	9·1	121·3	12·0	240]
43	6	PLL	7·15	97·6	97·6	390
40	6	PLL	5·6	79·2	79·2	317
43	19·6	0	2·33	40·0	40·0	3648
40	16·6	0	2·33	40·0	40·0	1824

Table 4 E.r.p.s required for the various systems

System	Modulation	A_r (m²)	T (deg K)	E.R.P. (kW)	Transmitted bandwidth (MHz)
Present domestic	a.m.	0·216	7000	1 200 000	
Above with dish antenna and tunnel diode amplifier	a.m.	0·786	350	16 500	
Dish antenna, tunnel diode amplifier and f.m. receiver	f.m.	0·786	350	84	50·0
Above with phase-locked loop	f.m.	0·786	350	25	79·2

Table 5 (B_3 = 12 kHz; tone/noise ratio 51 dB)

$10 \log \rho_t$ (dB)	$M^2(1+\overline{1-\beta M})$	β	M_1	B_1 (kHz)	B_2 (kHz)	B_0 (MHz)
13	2100	0	12·5	324	324	6·48
		½	15·5	396	218	4·36
		[1]	[45·8]	[1124]	[24]	[0·48]
10	1050	0	9·9	260	260	2·60
		½	12·2	314	177	1·77
		[1]	[32·4]	[825]	[24]	[0·24]
6		PLL	47	1128	1128	4·5

provided only that the receiver is adapted to accept the appropriate form of modulation. The ratios of vision carrier to sound carrier requirement are so large that one may infer that the sound carrier can be superposed on the vision carrier and that a single output stage can handle the combination with small effect on the required rating. Indeed several different sound channels, each with its own carrier are possible, a fact which may be valuable when the reception area is multi-lingual.

2. Satellite Aspects

Having established a basis for reckoning the r.f. power required for any given television service, it is now appropriate to explore the satellite implications.

Table 6 is an analysis of the relevant quantities over a range of satellite proposals. A remarkable conclusion emerges: the total satellite mass (in orbit) per watt of r.f. satellite power is substantially constant over the range (27 : 1) considered, and may be taken in round terms as 6·2 kg per watt of r.f. power.

(The only exception to this is the *A.T.S.* which shows the higher figure of 8·75 kg/W, but this is explained by the fact that *A.T.S.* is not specifically a communication satellite but carries much additional equipment.)

Another figure which emerges from Table 6 is that the ratio of raw d.c. power to r.f. power tends to be constant at around the value 5. As the satellite power is increased an extreme low limit of this would be the

Table 6

	Units	*Early Bird* category[5]	*Intelsat* 2 category[5]	*A.T.S.*	Hughes proposal 1965[3]	Ford Foundation Proposal
Powers:						
R.f. output	W	6	12	40	160	120
Total raw d.c. power	W	36	60	200	750	600
Masses:						
Power supply: solar cells, mounting, etc.	kg	3·9	8·6			91
Communications equipment	kg	12·2	12·2			70
Attitude and station control: hardware	kg	4·0	8·2			29
propellant	kg	4·5	9·5			110
Apogee motor: case and venturi	kg	4·5	9·4			52
propellant	kg	28·0	78·4			480
Structure and miscellaneous	kg	8·3	18·6			156
Total at launch	kg	67·6	154			988
Total in orbit	kg	39·7	75	350	1000	508
Costs:						
Satellite complete	$M	2·6	2·6			
Launching cost (including launcher cost)	$M	2·0?	2·5			
Total per orbited satellite	$M	4·6	5·1			
Ratios:						
Total raw d.c. power/r.f. power transmitted	—	6·0	5·0	5·0	4·7	5·0
Mass of power supply per watt raw d.c.	kg/W	0·108	0·143			0·151
Mass of power supply per watt r.f.	kg/W	0·65	0·72			0·755
Total mass into orbit per watt raw d.c.	kg/W	1·10	1·25	1·75	1·33	0·85
Total mass into orbit per watt r.f.	kg/W	6·61	6·25	8·75	6·25	4·24
Total cost per watt r.f. in orbit	$M/W	0·77	0·425			

reciprocal of the efficiency of the output stage, say 3. In practice for a large output we might expect to achieve perhaps 4.

For a television satellite we are likely to be interested in very high r.f. power. Can we look for a reduction in the specific mass? There seem to be only two areas where useful mass economy may be found, namely in the power supply mass and the station-keeping mass.

2.1. *Power Supply Mass: Solar Power*

All the satellites in Table 6 are solar-powered and the specific mass of the power supply is around 0·12 kg/W d.c. Great improvement is possible here, as shown by Treble.[6] He concludes that the most promising development is in terms of 'thin' (100 μm) n-on-p silicon solar cells mounted on 50 μm polyimide (Kaplon). He infers that:

specific mass	0·02 kg per watt of raw d.c. power
specific area	0·012 m^2 per watt of raw d.c. power
specific cost	$360 per watt of raw d.c. power

Unfortunately, although this represents a six-fold reduction in the specific mass of the power supply, the whole power supply contributes only some 10% of the total satellite mass, so that the improvement must reflect less than 10% on the overall specific mass.

Another question which arises is 'up to what power requirement are solar cell arrays practical?' In a weightless environment there does not appear to be any theoretical limit, and the question is one of packaging and deployment. The largest comparable array so far known to have been deployed in space is that of the NASA project *Pegasus* where an array of dimensions 29·3 m × 4·1 m was successfully deployed (1965). (Although this was not a solar cell array, but a meteorite detection array, this brilliant experiment yielded a result of interest to solar cell arrays, namely a puncturing meteorite rate of 1·3 per square metre per year!)

If we take this area of 120 m^2 as indicative of the present practical limit, this would imply a limit of 10 kW (raw d.c.) for solar power supplies.

Although there is nothing absolute about this limit, we must not lose sight of the mechanical difficulty presented by a satellite with one member stabilized relative to earth, and another (enormous) member stabilized to the Sun line, hence with a relative rotation of one revolution per day.

2.2. *Power Supply Mass: Nuclear Power Supplies*

It is rather easy to cite this as the solution of all satellite high-power problems. While this may or even must ultimately be correct, it is necessary to point out how far from practicality nuclear power supplies are at the present date.

Slone[7] gives an account of the NASA SNAP 8 development to August 1966. It is a splendid technical achievement and great credit must be given to those who have carried out the work, as also to the NASA authorities who have funded such a project 'in the abstract'. But whether it is applicable to a television satellite is doubtful as shown by the data given in Table 7.

Table 7

SNAP 8 data (1966)

Format	Nuclear reactor plus sodium-potassium heat-transfer fluid plus mercury turbine plus alternator
Power rating	35 kW electrical
Specific mass (estimated)	Present: 80 kg/kW Potential: 23 kg/kW
Life rating	1·2 years
Cost	Unknown

These data indicate that while SNAP 8 is nicely placed in power level, it does not presently compete with the solar power supply in terms of specific mass, though it may ultimately do so. For television satellite applications, however, a 10-year life is likely to be called for and the effect of this on (a) feasibility, (b) specific mass and (c) specific cost, is unknown.

2.3. *Station-keeping Mass*

Table 6 indicates that station- and attitude-keeping 'propellant' represents 12 to 25% of total in-orbit mass. Substantially the whole of this can be saved by going over to ion thrusters for attitude- and station-keeping.

2.4. *Total Potential Mass Saving*

In the broadest possible terms we may look for improvements of 10% and 25% of total satellite mass in respect of power supply and station-keeping. Thus we might expect 0·65 × 6·25, say 4 kg/W r.f. as a likely specific mass for a large satellite which can be reasonably envisioned as being developed.

2.5. *Launcher and Launching Costs*

In Fig. 2, an endeavour has been made to present specific launching costs ($k/kg) versus satellite mass in orbit. The solid curve is a mean through six specific American launchers, and it is somewhat surprising that the individual points lie so close to a smooth curve.[8]

Also plotted are European possibilities in terms of ELDO.[9] It is seen that ELDO costs are presently about twice the American costs, which is entirely reasonable in view of the great American advantage in experience. The ELDO study for 1975, however, suggests equality with the (present) American specific costs by that date.

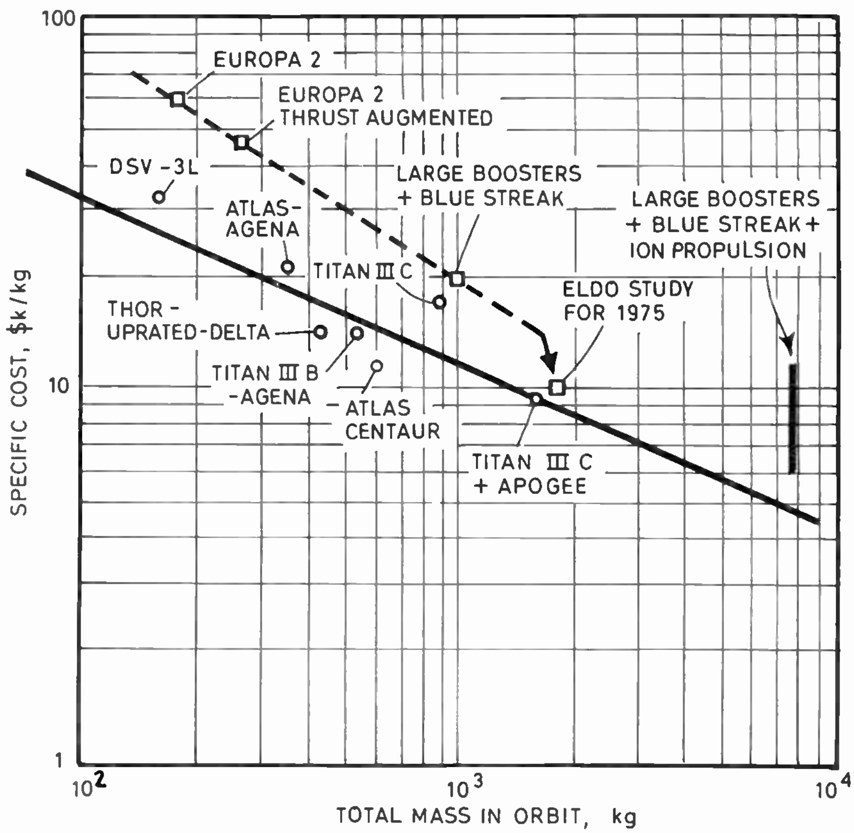


Fig. 2. Booster specific costs.

It is unfortunate that all these costs are to be interpreted only in the broadest terms. The basis on which development is amortized is obscure as is also the factual availability which is subject to political considerations. For lack of better information, in the next section we shall adopt the cost data indicated by the mean curve through the American data points.

A possible reduction of launching cost is envisaged in the concept of electrical propulsion for orbit transfer. The important proposal is made that a satellite requiring several kilowatts of power in orbit can well use this same power source for electrical propulsion during the transfer from low to geostationary orbit. The whole procedure has been ably propounded and analysed by Burt and colleagues.[10] Potential for economies appears to be great, but quantitative data are unfortunately not yet available.

2.6. *Total Satellite Costs*

An endeavour to present costs per watt of r.f. in orbit over a wide range of powers is to some extent stultified by the roughness of data on satellite costs. Indeed, this term has no absolute meaning as it is vastly dependent on the number of satellites to be produced and on the basis of development amortization. Subject to this disclaimer, Table 8 presents order of magnitude costs.

Table 8

R.F. power (W)	Satellite mass (kg)	Satellite cost ($M)	Launch cost† ($M)	Total satellite in orbit cost ($M)	Specific cost ($M/W) (r.f.)
12	75	2·6	2·8	5·4	0·45
72	464	5·0	7·4	12·4	0·18
160	1000‡	(7·8)§	11·5	19·3	0·12
500	2000‡	(11·5)§	17	28·5	0·057
1000	4000‡	(18·2)§	27	45·2	0·045
2000	8000‡	(24·3)§	36	60·3	0·030

† On basis of solid curve in Fig. 2.
‡ Masses computed from Section 2.5.
§ Assumed to be $\frac{5 \cdot 0}{7 \cdot 4}$ × launch cost.

3. Conclusions

The main conclusion from the above figures is that the concept of a direct television broadcasting satellite is far beyond the 'state of the art' so long as one stipulates the use of existing receivers (and therefore frequency assignments). However, as one recedes step by step from this stipulation, the proposal becomes more and more practical, and finally, if one admits new frequencies, new antennas and new or adapted receivers using wideband f.m., it becomes broadly within the 'state of the art'.

One may question, however, where exactly is the need for such systems? It would seem to be not in territories which are already high in programme distribution and receiver densities unless the issue becomes forced by extreme expansion in the demand for programme choice. On the other hand there are many areas where economic-cum-geographic considerations make conventional methods impossible or uneconomic. In such cases the relative economics of a satellite direct broadcast system come in for consideration. For such purposes the costs indicated in this paper constitute useful guidelines but require extensive refinement for ultimate factual costing.

It is to be noted, however, that the more the receiver departs from the present domestic article the more does the system become attractive for television distribution rather than direct broadcasting. Following this line of thought, the author is prompted to propose for consideration what might be called a 'tigon' system. The tigon, it seems, is a cross between two species, the unlikely but factual result of the (illicit) union between tiger and lioness; it is stated to exhibit the more vigorous characteristics of both species.

Gabriel[12] gives figures for the high technical and cost effectiveness of a wire distribution system in the form of city distribution centres feeding into individual home receivers by wire. The marriage of this with a satellite receiving centre is now an obvious step and, tigon-like, should be able to combine the advantages of the two extreme proposals—satellite and wired television.

4. References

1. Hilton, W. F., British Aircraft Corporation Ltd., Private communication, May 1967.
2. Matthei, W. G., 'Recent developments in solid state microwave devices', *Microwave J.*, **9**, pp. 39–46, March 1966.
3. Altovsky, V. A., 'Prospects in the Television Satellites Field', EUROSPACE, Brussels, October 1965.
4. Bray, W. J., 'Basic Communication Sciences', Discussion Meeting on Space Science and Technology, Royal Society, London, 27th February 1968.
5. British Aircraft Corporation Ltd., Classified study, August 1965.
6. Treble, F. C., 'Large Solar Cell Arrays', ELDO Symposium, Paris, 10th October 1967.
7. Slone, H. O., 'SNAP Development Status', NASA, August 1966.
8. Worden, D., Hughes Aircraft Company, Private communication, February 1968.
9. Pardoe, G. K. C., 'Launching Facilities and Payload Capabilities', Discussion Meeting on Space Science and Technology, Royal Society, London, 27th February 1968.
10. Burt, E. G. C., 'On space manoeuvres with continuous thrust', *Planet. Space Sci.*, **15**, pp. 103–22, 1967.

'Optimum Functions for Mass and Energy Flow in Space Manoeuvres'. R.A.E. Tech. Report 67267, 1967.

'Space Science and Electrical Propulsion', Discussion Meeting on Space Science and Technology, Royal Society, London, 27th February 1968.

11. Haggai, T., Hughes Aircraft Co. Internal Reports and U.S. Patent Nos. 3,346,814 and 3,346,815.
12. Gabriel, R. P., 'A comparison of wired and wireless broadcasting for the future', *J. Roy. Telev. Soc.* (To be published.)

5. Appendix

Derivation of Video and Audio Signal/Noise Ratios

Notation:

f_0 carrier frequency, undeviated

C carrier amplitude (volts)

$\pm f_d$ maximum frequency deviation due to signal

B_3 highest modulating frequency $\equiv$ information bandwidth

f single interfering signal frequency

E single interfering signal amplitude (volts)

For small values of E/C the interfering signal produces a carrier deviation $\pm(f-f_0)E/C$ at frequency $(f-f_0)$; hence a mean square frequency deviation

$$\tfrac{1}{2}(f-f_0)^2\frac{E^2}{C^2} = \tfrac{1}{2}(f-f_0)^2\frac{\overline{E^2}}{\overline{C^2}} = \tfrac{1}{2}(f-f_0)^2\frac{P_n}{P_c}$$

where $\overline{C^2}$ and $\overline{E^2}$ are mean square voltages of carrier and interfering signals and P_c and P_n are corresponding powers.

For the case where the interference is thermal in character, and has equivalent total noise temperature T, the noise power in frequency band df is $kT\,\mathrm{d}f$. Hence mean square frequency deviation due to noise in bandwidth B symmetrically disposed about f_0 is

$$\int_{f_0-\frac{B}{2}}^{f_0+\frac{B}{2}} \tfrac{1}{2}(f-f_0)^2\frac{kT}{P_c}\,\mathrm{d}f = \tfrac{1}{2}\frac{kT}{P_c}\left[\tfrac{1}{3}(f-f_0)^3\right]_{f_0-\frac{B}{2}}^{f_0+\frac{B}{2}}$$

$$= \tfrac{1}{2}\frac{1}{B_0}\cdot\frac{2}{3}\left(\frac{B}{2}\right)^3$$

$$= \frac{1}{24}\frac{1}{B_0}B^3 \qquad \ldots\ldots(15)$$

where $B_0 = \dfrac{P_c}{kT}$ the indicial bandwidth.

$$\left[\frac{\text{r.m.s. frequency deviation due to noise}}{\text{peak frequency deviation due to signal}}\right]^2$$

$$= \frac{1}{24}\cdot\frac{B^3}{B_0 f_d^2} \qquad \ldots\ldots(16)$$

Note that at the output of the i.f. amplifier, bandwidth B_2, the carrier/noise power-ratio is

$$\frac{P_c}{kTB_2} \quad \text{or simply} \quad \frac{B_0}{B_2}$$

This carrier/noise power-ratio at the discriminator we term the working threshold ρ_t.

For the effective signal/noise ratio at the output of the receiver we are concerned only with those interfering components which are effective in the final base frequency band; that is to say components in the band $f_0 - B_3$ to $f_0 + B_3$, a bandwidth $2B_3$.

Putting $B = 2B_3$ in (16) we obtain

$$\left[\frac{\text{r.m.s. frequency deviation due to effective noise}}{\text{peak frequency deviation due to signal}}\right]^2$$

$$= \frac{1}{24} \cdot \frac{8B_3^3}{B_0 f_d^2}$$

$$= \frac{1}{3} \frac{B_3^3}{B_0 f_d^2}$$

$$= \frac{1}{3} \frac{B_3^3}{B_0 M^2 B_3^2}$$

$$= \frac{1}{3} \frac{B_3}{B_0} \cdot \frac{1}{M^2} \qquad \text{......(17)}$$

or, inverting,

$$\left[\frac{\text{peak frequency deviation due to signal}}{\text{r.m.s. frequency deviation due to effective noise}}\right]^2$$

$$= 3 \frac{B_0}{B_3} M^2 \qquad \text{......(18)}$$

$$= 3 \frac{B_0}{B_2} \cdot \frac{B_2}{B_3} M^2$$

$$= 3\rho_t \frac{B_2}{B_3} M^2$$

$$= 6\rho_t M^2(1+M) \qquad \text{......(19)}$$

The validity of these expressions is subject to the value of ρ_t being sufficiently high to justify the assumption that output signal is conveyed only by frequency variation, the effect of amplitude variation being suppressed. This implies a minimum value for $10 \log \rho_t$ of 13 dB for a normally good discriminator/limiter system, or 10 dB for exceptionally well-designed circuitry.

All the above refers to the case of a straightforward f.m. receiver without f.m. feedback. To introduce the latter important principle let the feedback be such that a deviation f_d is reduced to $(1-\beta)f_d$. The necessary i.f. bandwidth B_2 may now be reduced to

$$2(B_3 + \overline{1-\beta} f_d) = 2[B_3 + (1-\beta)MB_3]$$

or

$$B_2' = 2B_3[1+(1-\beta)M] \qquad \text{......(20)}$$

This bandwidth B_2' is appropriate to the evaluation of the working threshold. However, since f.m. feedback affects signal and noise voltages alike, it does not otherwise affect the output signal/noise ratio.

(To elucidate this point further, suppose that we consider the introduction of the f.m. feedback feature in two stages:

(a) Application of f.m. feedback without reduction of i.f. bandwidth.

(b) Reduction of i.f. bandwidth.

At stage (a) no change of signal/noise ratio results.

At stage (b) the components eliminated are not those directly contributing to observable noise in the output but only those contributing to the total i.f. noise applied to the discriminator.)

From eqn. (18) we have

$$\left[\frac{\text{peak frequency deviation due to signal}}{\text{r.m.s. frequency deviation due to noise}}\right]^2 = 3\frac{B_0}{B_3} M^2$$

$$= 3 \frac{B_0}{B_2'} \frac{B_2'}{B_3} M^2$$

$$= 6\rho_t M^2[1+(1-\beta)M] \qquad \text{......(21)}$$

For the television case, from the point of view of picture signal/noise ratio we are concerned with that portion of the total frequency deviation ($2f_d$) corresponding to a picture range of black to white. With normal standards this amounts to $1{\cdot}41 f_d$, hence

$$\left[\frac{\text{frequency deviation range corresponding to picture range black to white}}{\text{r.m.s. frequency deviation due to noise}}\right]^2$$

$$= \left[\frac{1{\cdot}41 f_d}{\text{r.m.s. frequency deviation due to noise}}\right]^2$$

$$\simeq 12\rho_t M^2[1+(1-\beta)M] \qquad \text{......(22)}$$

For the sound channel, it is normal to work in terms of a single frequency test tone which fully modulates the system. Equation (21) becomes

$$\left[\frac{\sqrt{2} \times \text{r.m.s. deviation due to test tone}}{\text{r.m.s. deviation due to noise}}\right]^2$$

$$= 6\rho_t M^2[1+(1-\beta)M]$$

$$\left[\frac{\text{r.m.s. deviation due to test tone}}{\text{r.m.s. deviation due to noise}}\right]^2$$

$$= 3\rho_t M^2[1+(1-\beta)M] \qquad \text{......(23)}$$

Manuscript received by the Institution on 19th July 1968. (Paper No. 1217/Com 7.)

Electronics and Aircraft

It has often been stated, perhaps in jest, that modern military aircraft are merely high-speed vehicles for electronic equipment. The point was however given considerable support at the recent Exhibition and Flying Display held by the Society of British Aerospace Companies at Farnborough, when the Hawker Siddeley *Nimrod* was demonstrated (in flight). This is a new maritime reconnaissance anti-submarine aircraft, based on the *Comet* airliner design, which is to enter service with the Royal Air Force next year. Details of its complement of electronic equipment were announced shortly before the Exhibition.

Nimrod is equipped with the latest operational airborne sensor devices for detecting, identifying and locating surface or submerged vessels and has provision for new systems still in the experimental stage. A.s.v. radar is used to detect surface shipping or submarine periscopes and snorkels and the scanner located in the nose radome can also be used for cloud collision warning. Electronic counter measures (e.c.m.) equipment, using a fin-top aerial, detects and locates sources of radar transmission and the exhaust trail indicator traces diesel exhaust gases in the atmosphere.

The sonar system gives accurate short range location of an underwater target using active or passive buoys, which can locate stationary or moving submarines and provide range and bearing data. An 'on-top' position-indicator in the aircraft gives precise indication of passage over operating sonobuoys, while the magnetic anomaly detector (m.a.d.) mounted in the tail-boom indicates the presence of submerged vessels from their disturbance of the Earth's magnetic field.

The internal layout of *Nimrod* includes a central operations room which the Royal Air Force considers will enhance crew efficiency. Here the sensor input data, continuously processed by an Elliott MCS 920B digital computer, is displayed at a combined console manned by the tactical navigator and the routine navigator. The tactical navigator has a 24-in diameter cathode-ray tube giving a continuous, North-stabilized picture of the tactical situation showing aircraft position and present and past track, sensor inputs and sonobuoy positions.

Bearing lines and range circles from the sonobuoys, e.c.m. bearings, m.a.d. and visual and radar fixes, are also represented on the display which covers 10 to 160 nautical miles depending on which of the five scales is selected. A smaller rectangular display gives related alpha-numeric information, but can take over the situation display should the larger tube fail. Aircraft parameters, target track and speed, sonobuoy channel numbers and symbol lat/long co-ordinates are displayed.

The computer in the *Nimrod* system predicts target position, track and speed from sonar active and passive fixes. Predicted intercept point, and weapon splash and release points, are continuously computed and updated as new fixes are obtained and as the aircraft position changes.

The routine navigator sits at a combined console with the tactical navigator, and controls the long-term navigation of the aircraft, e.g. to and from the search area. The primary mode of navigation on *Nimrod* uses accurate inertial heading information from the Elliott E3 stable platform combined with drift and ground speed from the Decca Doppler after correction for sea motion. During tactical manoeuvres when the Doppler unlocks, inertial velocities are used to give ground speed, and the Doppler aerial is slaved to the correct drift angle based on the track derived from inertial velocities and true heading to ensure speed relock when the manoeuvre is complete.

To supply processed information to the autopilot, navtac system, altimeters, machmeters, etc., an Elliott monitored air-data computer is installed. Radio height is provided by twin radio altimeters, and is used with barometric height from the air-data computer to give a high integrity radio/barometric height lock for the automatic flight control system.

The radio operator's station has 1 kW airborne equipment and offers immediate world-wide communications on any channel in the h.f. band from 2 to 30 MHz.

Ground-based aircraft aids prominently displayed at Farnborough included a Board of Trade exhibit of the National Air Traffic Control Services' *Mediator* plan.

Mediator is the code name given to the comprehensive plan to provide a computerized radar-based air traffic service to all users of the airspace above 5000 ft over the U.K. Flight Information Region. For this purpose one of the largest computer complexes used in air traffic control anywhere in the world is presently being designed and installed in a new London Air Traffic Control Centre (LATCC) at West Drayton, Middlesex. There will be a smaller, but still large, complex at Prestwick, Ayrshire.

The focal point of the stand was an air traffic control sector suite typical of those to be used at West Drayton under Stage 1 of the plan. Other exhibits included models of a typical N.A.T.C.S. radar/radio station and long-range surveillance primary and secondary radar heads. Primary and secondary radar systems are being provided at seven sites in the area.

Co-operation with the air traffic control service will be mandatory in all controlled airspace airways and within the 164,000 square miles of the upper airspace control area. Below 25,000 ft and outside controlled airspace, in the middle airspace, participation will be optional but it is expected that 90% of all aircraft flying above 5,000 ft will make use of the service during part of their flight.

Stage 1 of the *Mediator* plan is mainly concerned with the introduction of improved facilities and equipments and the exploitation of their individual capabilities in the immediate future. Stage 2 of the plan concentrates on improvements to the facilities and equipments and progressively more advanced techniques and integration in order to exploit their collective capabilities to the full. In particular more sophisticated data processing facilities to be provided under Stage 2 will relieve the air traffic controllers of many routine tasks. When Stage 2 is reached, LATCC will have control of all the southern part of the United Kingdom airspace as far North as 55° for controlled airspace, 56°N for middle airspace and 57°N for upper airspace.

U.D.C. 681.32

A Brief Review of Computer Assessment Methods

By

R. F. WICKENS,

C.Eng., M.I.E.E., M.I.E.R.E.†

Presented at the Joint I.E.R.E.-I.E.E. Computer Groups' Discussion Meeting held in London on 2nd April 1968.

Summary: The existing methods of assessing computer performance are reviewed. The advantages and limitations of each method are discussed. Finally, possible future areas of study are indicated.

1. Introduction

If the assessment of computer performance is not to be merely a matter of opinion it must be interpreted as the measurement of computer performance. This paper will give a brief review of the methods currently available and indicate their limitations and finally suggest areas for future study.

Before going on to discuss the methods available we should first consider why we want to measure performance and the uses to which the results will be put.

Users will clearly wish to determine the 'best buy' in their circumstances, estimate job running times, optimize their configurations and estimate future machine requirements.

The computer manufacturers will require to compare their products with those of their competitors, to plan future products and to optimize the performance of these products.

Computer performance can be judged by a variety of criteria but these fall into three basic categories:

(a) economic criteria such as throughput per pound sterling;

(b) utilization criteria such as throughput per shift;

(c) operational criteria such as ease of use, staffing levels and reliability.

The weight to be given to any particular category of criteria, of course, depends upon circumstances.

2. Factors Affecting Computer Performance

The performance of a computer is the combined performance of its hardware and software: and this performance must in general be measured in relation to the work to be done. Ideally, we should take into account the 'efficiency' of the computer system's throughput and measure not the work performed but only the effective work carried out on the user's problems. However, the farther away from the hardware one moves the more difficult it becomes to assess performance.

The performance of a computer is influenced by a large number of factors. For example, on the hardware side performance will be influenced by the speed at which instructions are obeyed, word length, store size, 'power' of the order code and the number and characteristics of the peripherals to indicate but a few. On the software side performance will be affected by such things as the facilities offered by the software, the computer resources that have to be dedicated to it and the 'efficiency' with which the software provides its facilities.

3. Methods of Measuring Performance

Essentially there are two approaches to the problem of measuring the performance of a computing system: the analytical and the experimental. A system can be analysed in terms of parameters describing aspects of its operation or its constituent parts and from these parameters can be derived a measure of its performance. Alternatively a 'live' or 'paper' experiment can be carried out on a system and from the results the performance may be deduced.

3.1. *Some Simple Methods*

One of the simplest methods of comparison is to compare the cycle time of the main storage. However, as machines can have stores with 'widths' varying from one character to several words some misleading conclusions can easily be drawn. A considerable improvement over this method is to compare the number of bits accessed per second. In as much as this method does tell one something about the rate at which data can be handled it is of some limited value.

Another fairly common method is to compare individual instructions times such as 'add' or 'multiply'. Whilst such a scheme can be of use when comparing machines with a common architecture its use when comparing dissimilar machines can be very misleading.

† Ministry of Technology, Computer Advisory Service, Technical Support Unit, London, E.C.1.

3.2. *Instruction Mixes*

One aspect of computing power is the speed with which instructions are obeyed. A popular method of measuring this is to use instruction mixes. In effect an 'average' instruction is defined for the intended work load and it is then determined how fast these 'average' instructions are obeyed. The frequency distribution of the instruction types within a mix can be obtained in a variety ways including manual analysis of programs, hardware or software monitoring of a machine during live work or sampling a machine's activity at, say, 1 second intervals. The results of a mix calculation are expressed either in terms of instructions obeyed per second or the time to obey a unit of work made up in accordance with the mix.

The use of instruction mixes does not solve the problem of comparing machines with different architectures and introduces the doubtful concept of the make-up of an 'average' program for a given type of work. Also a mix figure tells one nothing about other important aspects of computing power such as input–output capability, store size, effectiveness of software, etc. Despite the seeming simplicity of an instruction mix, in order to ensure consistent results mixes must be rigorously defined and applied in a consistent manner.

Nevertheless, despite their failings mix calculations are widely used. Provided they are interpreted and used with considerable caution they can give a crude measure of computing power.

3.3. *The Post Office Work Unit*

The Post Office Work Unit is another commonly used measure of computer performance. It is in effect a high-level mix—a mix defined in machine-independent terms. The result is expressed as the time to carry out the defined unit of work. Graphs are plotted of the number of characters transferred per second from tape buffers to the core store against the number of P.O. Work Units performed for each 1000 characters transferred. The area under the graph between realistic activity limits is taken as a measure of the merit of a system.

It is a fairly lengthy process calculating a P.O. Work Unit and care must be taken to ensure consistent application to all systems. Also the Unit was designed specifically with P.O. work in mind but a user can define his own work unit. Nevertheless, the method is a useful evaluation tool and as used by the G.P.O. does take into account the effects of input–output. However, when used merely to express computational speed, as it often is, it is not very much better than a mix figure, although, because of its machine-independent definition, it is perhaps better for comparing machines with different architectures.

3.4. *Standard Routines and Benchmarks*

Two popular methods that are examples of the experimental approach are the use of standard routines and benchmarks. Standard routines are important or representative elements of the work load such as a file up-date or mathematical routine. The technique relies on the machine's performance of the routine being related to that on live work. Timing of such routines can provide some idea of a machine's performance on a particular type of work but it is difficult to choose or design appropriate routines and to judge system throughput on a real work load from the results.

An extension of the standard routine idea is the use of benchmark problems. These consist of short but complete standard problems representative of the type of work to be done. The benchmarks are coded for each of the machines of interest and their running time calculated. Users can devise their own benchmarks or use one of the well-known ones such as those defined by the Auerbach Corporation.

Benchmarks have the advantage that they are defined in a machine-independent manner and the resultant timings can take into account all hardware aspects of the system. However, the calculation of timings is tedious and error prone and the effects of software are excluded.

A further extension of the benchmark method is actually to program the benchmarks in a high level language and then run them on the machines of interest. The resulting timings should then provide much valuable information on machine throughput and the state and effectiveness of the software.

3.5. *Simulation*

Yet another experimental technique is the use of simulation. It can be used by manufacturers to evaluate and optimize a new design without the construction of costly hardware. Also software can be written, tested and evaluated without waiting for the hardware to be constructed. Such techniques can be of very great value but the difficulties of simulating a complex computer system should not be underestimated.

Simulation has been used to aid the selection of computers. Available packages construct generalized models of computers and their work loads and these models are supplied with the appropriate parameters for each system of interest to the user. The package will provide estimates of job running time, core usage and so on. Because of the complexity of computer systems the models are usually considerable simplifications of the real world. Nevertheless, they can provide useful comparative information but the results must be interpreted with an understanding of the method and its limitations.

4. The Evaluation of Software

So far methods of assessing the hardware have been considered but software is of considerable and growing importance. Any measure of a machine's performance that does not take into account the effects of software must be highly suspect. Unfortunately few formalized methods appear to be available. The question that must be answered is 'what effect will the software have on the throughput of an installation?' One can attempt to answer this question two ways: by a detailed study of the software or by experiment. Of course what is usually measured by an experiment is the combined effectiveness of the hardware and software.

In effect when assessing software one has to determine the advantages to be gained from using it and weigh these against the price to be paid in terms of time and computer resources. The software can be considered in four parts:

(a) the essential software such as executive or supervisory routines;

(b) the operating system;

(c) the compilers;

(d) the applications packages.

The most effective methods of assessing software appear to be those relying upon live demonstrations of the efficiency of the facilities offered. However, the following should be investigated:

(i) the specification of the software;

(ii) the quality of the documentation including such items as manuals, flow charts and program listings;

(iii) the machine resources required by the software;

(iv) the availability of the software;

(v) the 'reliability' of the software.

5. The Effect of the Work Load

The throughput of a system can be profoundly influenced by its work load and any assessment method must take it into account. A major problem when defining a work load is in deciding what are the important parameters. Most techniques demand the use of arbitrary or average values for the chosen parameters rather than use their real distribution.

Another aspect of the work load is the effectiveness of the users' programs. It is unlikely that these programs will be examples of model programming. Good software is of value here in that it relieves the programmer of some of the difficult problems.

6. Conclusions

There is a variety of assessment methods available but most are essentially measurements of hardware capability. The mix and benchmark appear to be the most popular methods. The mix and allied methods consider but one aspect of hardware performance but are useful guides when the work load is computer bound.

Benchmarks can provide a valuable tool for computer evaluation but they must be properly defined. It will often be found that the results will correlate with some simpler criteria such as a mix or peripheral transfer rate. If meaningful results are to be obtained more than one benchmark should be used in order to counteract any accidental or deliberate bias in the choice of the benchmarks. An attractive technique is the use of a mix of benchmark programs that represent the expected work load. If the most is to be extracted from the use of benchmarks they should be actually programmed and run on the machines of interest.

As yet there appears to be no completely satisfactory computer assessment method and computer assessment is currently still something of an art. Users should not rely on any single evaluation method but should use those methods that appear to be the most appropriate in their circumstances.

Much work remains to be done on the development of computer assessment methods, especially methods that include the influence of software. A method or methods are required that are applicable to the whole spectrum of computers from small process control machines to large networks of computers that take fully into account the environment in which they operate. Progress is slow because of the lack of fundamental data from which new methods may be developed. Much of these data reside within existing installations and perhaps the first thing to do is to collect this information and attempt to digest it. If this is to be done in a reasonable time-scale more co-operation between all interested parties will be required.

The computer art is advancing rapidly. Future systems may well be very different from current designs. Future assessment methods must be able to cope with multi-processor, multi-programming and multi-access systems perhaps connected into computer networks to provide fail-soft facilities. A useful analogy for a large computer system might be the telecommunication system where such concepts as grade of service are of value.

Manuscript received by the Institution on 13th May 1968. (Contribution No. 108/C106.)

U.D.C. 681:32

Assessing Computer Performance

By

J. MEREDITH SMITH,
B.Sc., Ph.D.†

Presented at the Joint I.E.R.E.-I.E.E. Computer Groups' Discussion Meeting held in London on 2nd April 1968.

Summary: The dynamic behaviour of a computer system is analysed by representing it as a network of queues and phases which are then treated as a Markov chain.

For the purposes of this discussion it is convenient to consider a job, which is being undertaken by a computer, as consisting of a number of tasks or phases (e.g. input, loading, compiling, computing, etc.). Modern computers are capable of handling a number of these phases simultaneously and hence a computer may hold at any instant a number of jobs at various stages of completion. Such a complex of jobs and facilities for job execution we shall call a system, and it is the performance of the system carrying out its job load that we wish to determine. The method we shall describe is in contrast with the more well-known techniques of computer performance evaluation such as mixes (which assess the performance of a single computing element) and bench-marks (which assess the performance of the system in carrying out an isolated job).

Let us consider the situation within the system at some instant in time, t. We label the phases of the system by

$$i = 1, 2, \ldots, N$$

where there are N possible phases. At time t we shall suppose that there are $n_i(t)$ jobs in phase i and $q_i(t)$ jobs which have completed phase j, $(j < i)$ and are queuing to enter phase i. Hence we obtain a $2N$ dimensional vector,

$$\mathbf{x}(t) = \{q_1(t), n_1(t), q_2(t), n_2(t), \ldots, q_N(t), n_N(t)\}$$

whose value at time t we shall call the state of the system at time t. To analyse system performance we have to determine the value of this vector.

First we note that if $n_i(t) = 0$, then $q_i(t) = 0$, since if a phase is not occupied there can be no queue awaiting entry to it. In the case of some phases only one job may occupy the phase at a time (e.g. in a normal system only one job is being executed by the c.p.u. at any instant although others may be queuing for service); in these cases n_i can take the values 0 and 1 only. In many cases there are also constraints imposed on the values of q_i; e.g. there may be a maximum permitted number of jobs in the system

$$\sum_{i=1}^{N} q_i < K$$

or individual maxima

$$q_i < m, q_i + q_j < l.$$

These constraints characterize the system. As a job passes from a queue to a phase or a phase to a queue the system state changes. If there are M possible states of the system

$$\mathbf{x}_r, r = 1, 2, \ldots, M$$

we can define operators Q_{rs} which cause the system to change from $\mathbf{x}_r$ to $\mathbf{x}_s$. Furthermore it is possible to have phases which do not possess a prior queue (this is necessary when modelling non-exponential distributions—see below) and in these cases there will be constraints imposed upon the preceding phase. It is also possible for the system to contain branches and loops with fixed probabilities of jobs taking the branches.

In analysing the dynamic behaviour of such systems we can use the technique of representing the system as a network of queues and phases. We shall assume that the rate of arrival of jobs to the system is Poisson with a mean rate λ_0 and that the time distribution $F(t)$ of occupancy of the phases by jobs is negative exponential, i.e.

$$F(t) = 1 - e^{-\lambda_i t} \qquad \ldots\ldots(1)$$

where $1/\lambda_i$ is the mean time of occupancy of the phase i.

Under these conditions the system can be represented by a Markov chain.[1] If we now define the probability vector

$$\mathbf{P}(t) = \{P_r(t)\}, \quad r = 1, 2, \ldots, M \qquad \ldots\ldots(2)$$

where $P_r(t)$ is the probability that the system is in the state r at time t, we find that

$$\dot{\mathbf{P}}(t) = \mathbf{A}\mathbf{P}(t) \qquad \ldots\ldots(3)$$

where $\mathbf{A}$ is the matrix of transition rates, the elements a_{rs} being the values of λ_i which cause a change of state from r to s. For a distribution of the form of eqn. (1) the probability of a transition in a time δt is $\lambda_i \delta t$.

† English Electric Computers Ltd., Park Royal, London, N.W.10. (now part of International Computers Ltd.)

Hence from eqn. (3)

$$\mathbf{P}(t)+\dot{\mathbf{P}}(t)\delta t = \mathbf{AP}(t)\delta t+\mathbf{IP}(t) \qquad \ldots\ldots(4)$$

and if δt is sufficiently small that only one transition occurs during δt then

$$\mathbf{P}(t+\delta t) = (\mathbf{A}\,\delta t+\mathbf{I})\mathbf{P}(t)$$
$$= \mathbf{UP}(t), \text{ say} \qquad \ldots\ldots(5)$$

The non-zero elements of **U** are now of the form $\lambda_i \delta t$ and represent the probability of a transition from state r to s. In the steady state eqn. (5) reduces to

$$\mathbf{P} = \mathbf{UP} \qquad \ldots\ldots(6)$$

which can be solved iteratively on a computer; the elements of **P** give the 'long run' probabilities of queue lengths and processor occupation from which through-put may be determined. Further details of this technique are expounded elsewhere.[2, 3]

We shall illustrate this technique with a simple example; further examples are given by Smith.[3] Suppose that programs are stored in some backing store such as a disk and that users make requests from consoles for their use. On a request being made the job is placed in a queue for loading, loaded into core when at the top of the queue and then executed. For simplicity we shall assume that all distributions are negative exponential, that no more than one job can be in core, and that no more than two jobs can be queuing. The possible states are then given by

	q	n_1	n_2
x_1	0	0	0
x_2	0	0	1
x_3	0	1	0
x_4	1	0	1
x_5	1	1	0
x_6	2	0	1
x_7	2	1	1

In this case, owing to the absence of a queue prior to n_2, the constraints on the system are that $n_1+n_2 \leqslant 1$ and that if $n_1+n_2 = 0$ then $q = 0$. Diagrammatically the system can be represented as in Figure 1. Here λ_0 is the mean arrival rate, λ_1 the mean occupancy time of loading and λ_2 the mean occupancy time of execution. In this case we find that

$$\mathbf{A} = \begin{bmatrix} -\lambda_2 & 0 & \lambda_0 & 0 & 0 & 0 & 0 \\ \lambda_2 & -\lambda_1 & 0 & \lambda_0 & 0 & 0 & 0 \\ 0 & \lambda_1 & -\lambda_2-\lambda_0 & 0 & \lambda_0 & 0 & 0 \\ 0 & 0 & \lambda_2 & -\lambda_1-\lambda_0 & 0 & \lambda_0 & 0 \\ 0 & 0 & 0 & \lambda_1 & -\lambda_2-\lambda_0 & 0 & \lambda_0 \\ 0 & 0 & 0 & 0 & \lambda_2 & -\lambda_1-\lambda_0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_1 & -\lambda_0 \end{bmatrix} \qquad \ldots\ldots(7)$$

from which we can determine **U** and evaluate the steady state value of **P**.

This technique can be used to model systems with non-exponential distributions if these distributions can be expressed as combinations of exponentials; e.g. Erlang or hyperexponential (see, for example, Ref. 2). Experience from certain installations[4-6] leads us to believe that in many cases such Erlang and hyper-exponential distributions are good approximations to the observed distributions. The above example, which is trivial, generated a 7×7 matrix; it is clear that in dealing with large systems very large matrices are required. However, due to the fact that the matrix is sparse and due also to the repetitive property exhibited by the non-zero elements, techniques have been developed[2, 7] for the compact storage of these matrices on a digital computer.

This technique for analysing system performance has been used successfully in the analysis of multi-access systems[8] and gives us estimates of response time, throughput, utilization of hardware and expected queue lengths. It has also been used to analyse operating systems[9] and is a possible technique for analysing the behaviour of processor systems, multi-processors, communication systems and compilers.

The time taken to undertake a job on a modern computer is highly dependent not only on the hardware and software, but also on the nature and number of other jobs within the computer. This technique enables us to make some estimate of the performance of such a situation.

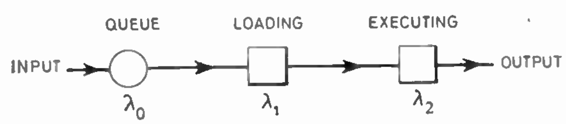


Fig. 1. Diagrammatic representation of the system.

References

1. Feller, W., 'An Introduction to Probability Theory and its Applications', Vol. 1 (Wiley, New York, 1957).
2. Wallace, V. L., and Rosenberg, R. S., 'Markov models and numerical analysis of computer systems', Proceedings of the Spring Joint Computer Conference, pp. 141–8, 1966.
3. Smith, J. L., 'An analysis of time-shared computer systems using Markov models', *ibid.*, pp. 87–95.
4. Walter, E. S. and Wallace, V. L., 'Further analysis of a computer centre environment', *Commun. Assoc. Computing Machinery*, **10**, pp. 266–72, May 1967.
5. Symons, C. R., 'A study of the workload and throughput of the CDC 6600 with SIPROS', CERN-DD/CO/67/2, January 1967.
6. Wood, B. M. and Koenigsberger, J. G., Private communication.
7. Wood, B. M. and Siday, A. P., Private communication.
8. Scherr, A. L., 'An Analysis of Time-sharing Computer Systems', U.S. Government Research Report AD 470–715 (MIT MAC-TR-18) 1965.
9. Foley, J. D., 'A Markovian model of the University of Michigan executive system', *Commun. Assoc. Computing Machinery*, **10**, pp. 584–8, September 1967.

Manuscript first received by the Institution on 2nd April 1968 and in final form on 11th June 1968. (Contribution No. 109/C107.)

STANDARD FREQUENCY TRANSMISSIONS

(Communication from the National Physical Laboratory)

Deviations, in parts in 10^{10}, from nominal frequency for **October 1968**

October 1968	24-hour mean centred on 0300 U.T.			October 1968	24-hour mean centred on 0300 U.T.		
	GBR 16 kHz	MSF 60 kHz	Droitwich 200 kHz		GBR 16 kHz	MSF 60 kHz	Droitwich 200 kHz
1	—	− 0·1	− 0·3	**17**	− 300·0	0	0
2	− 300·2	0	− 0·4	**18**	− 300·1	0	0
3	− 299·9	0	0	**19**	− 300·0	0	0
4	− 300·2	0	0	**20**	− 300·0	− 0·2	0
5	—	0	0	**21**	− 299·9	0	0
6	—	0	0	**22**	− 300·0	—	0
7	—	− 0·1	0	**23**	− 300·1	0	0
8	− 299·9	0	− 0·1	**24**	− 300·1	− 0·1	0
9	− 300·0	0	− 0·1	**25**	− 300·1	− 0·1	− 0·1
10	− 300·0	− 0·1	− 0·1	**26**	− 300·1	− 0·1	− 0·1
11	—	− 0·1	− 0·1	**27**	− 300·2	0	− 0·1
12	− 300·0	0	− 0·1	**28**	− 300·0	0	− 0·1
13	− 300·0	0	− 0·1	**29**	− 300·1	0	− 0·1
14	− 299·9	0	− 0·1	**30**	− 300·0	− 0·1	0
15	− 299·9	0	− 0·1	**31**	− 299·9	− 0.1	− 0·1
16	− 300·0	− 0·1	0				

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

U.D.C. 621.391.822:621.373.544

Noise Susceptibility of Integrated Circuits in Digital Systems

By

W. J. HOSIER†

Summary: The particular effects of noise upon logic gate and flip-flop circuits and the resultant general effects upon digital systems are considered. Common sources of noise are outlined and their significance in integrated circuit systems is considered. Such sources include crosstalk, reflections, common-earth impedance coupling and supply line noise.

Present methods of assessment and specification of d.c. noise margin and a.c. or pulse noise immunity are reviewed and examples are given of the application of such methods to representative commercial integrated circuits.

1. Introduction

The basic elements of a digital system are logic gates, flip-flops (bi-stable elements) and other storage devices. All of these are two-state devices, since they have a definite state associated with each of the system's two logic levels. In the case of logic gates, these states depend upon the persistence of the input levels. Each electronic device should be in one or other of its two defined output states, hereafter referred to as HIGH and LOW. This terminology avoids the restrictions of a particular logic convention.

Transition of a logic gate's output from one state to the other is initiated by a change of level at the gate input. However, the change in input level may be due to legitimate logic signals or due to noise. In the latter case a false transition of the gate's output may occur and the probability of the occurrence of a false transition is a function of the noise waveform and the circuit's response. Both the d.c. and transient response of the circuits are important and both of these aspects will be considered in the paper.

One of the most common contemporary applications of digital circuits is the computer and here designers find it essential to consider noise problems.[1–3]

Although false operation of logic gates is important, the false operation of flip-flops is particularly so, since once a flip-flop has been triggered into one of its stable states by noise, it will remain in this state even after the noise has decayed. The information represented by the new state of the flip-flop will then be propagated through the system as legitimate information. The seriousness of the false operation of system elements will depend upon the form of the system, therefore it is impossible to generalize in this matter.

2. Integrated Circuits

Of the three basic types of integrated circuit (thin-film, hybrid and silicon monolithic), the silicon monolithic has the widest usage. Digital circuits lend themselves particularly well to monolithic integration.

Since both active and passive components are formed within a small silicon chip, the permissible circuit power dissipation is limited. Furthermore, the initial areas of application of digital integrated circuits were the aerospace and military fields, which demanded minimal power dissipation and low-power supplies. Therefore, most present-day monolithic circuits have low-supply voltages, usually in the range 3 V to 6 V. This results in a small logic swing, making integrated circuits more susceptible to noise than many discrete component circuits. Typical values of supply voltage and logic swing for commercially available integrated circuits are shown in Table 1 in the Appendix.

As circuit speed increases (i.e. shorter transition times) the noise generated by the circuits also increases in magnitude. At present, non-saturating integrated logic circuits, such as ECL, can achieve transition times of 3 ns and propagation delays of 5 ns. Since discrete component circuits have not been developed to approach these speeds, the noise generated, due to circuit switching, in systems using integrated circuits may exceed that generated in systems using discrete component circuits. In addition, integrated circuits, due to their shorter transition times, can respond to noise pulses which would be completely rejected by discrete component circuits. As circuit speed increases, noise considerations become more important, since the noise generation capability of circuits is enhanced while the noise rejection capability is diminished.

† Formerly at the Department of Science and Electrical Engineering, College of Technology, Letchworth, Hertfordshire; now with British Aircraft Corporation, Stevenage, Hertfordshire.

3. Sources of Noise

3.1. *Crosstalk*

The study of crosstalk has a long history, many of the classical papers published on this topic having been concerned primarily with open-wire crosstalk, particularly in telephone systems.

Crosstalk may be defined as the induction of signals in a line or circuit due to the action of the stray magnetic and electric fields from a nearby, physically isolated line or circuit. The source of the stray magnetic and electric fields may be external to the digital system, in which case adequately designed screening can prevent this field from coupling with the interconnecting lines of the digital system. Within the digital system, the physical organization can be such that the electronic logic system as a whole is screened from any noisy devices. Therefore the main cause of crosstalk which remains is the switching action of the logic circuits themselves. This situation is usually represented as in Fig. 1.

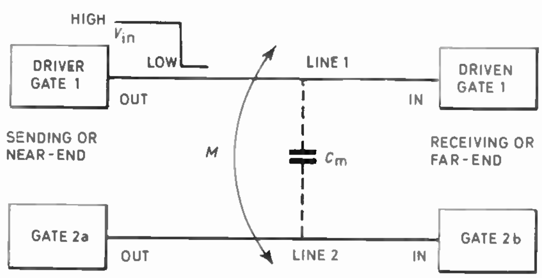


Fig. 1. Crosstalk between interconnections.

Line 1 is usually referred to as the active or disturbing line and line 2 as the quiet or disturbed line. When a switching transition occurs at the output of driver gate 1, as shown in Fig. 1, energy is transferred from line 1 to line 2 by both magnetic field coupling and electric field coupling. This is represented diagrammatically by the mutual inductive coupling M and the mutual capacitive coupling C_m. Due to the wide range of types of interconnection and the wide variation in length of interconnections in use in modern digital systems, there is a need for crosstalk analysis in terms of lumped parameters and distributed parameters, for both homogeneous and non-homogeneous media.

Analysis of crosstalk in terms of distributed parameters, is carried out by Gray[4] and Feller, Kaupp and Di Giacomo[5] using the conventional transmission line equations and by Jarvis[6] working from first principles, using the differential equations of the system. The expression for the crosstalk voltage produced on line 2 is similar in all three papers. It is shown to comprise three waves, one forward travelling (i.e. towards far end), one backward travelling, and one so-called differentiated wave. Of the three, the differentiated wave is the most important, since its amplitude is a function of the distance travelled along the line and the rate of change with time of the driving voltage on line 1. Therefore, for large logic swings, short signal transition times and physically long lines, the amplitude of this wave may be significant.

Oscillograms of near-end and far-end crosstalk waveforms for a specific transmission configuration, shown in Fig. 2, can be obtained[7] under various conditions of line termination. The far-end crosstalk is extremely important since this is the waveform seen by the input of gate 2b and may be responsible for false operation of this gate. (Fig. 3.)

Garth and Catt[8] also analyse crosstalk in microstrip transmission systems in qualitative terms. They show that the forward travelling wave is composed of the odd-mode signal, which propagates in differential mode between lines 1 and 2, and the even-mode signal, which propagates in common-mode with lines 1 and 2 as one conductor and the earth plane as the other. This theory is entirely compatible with the equations derived by Gray,[4] Freller *et al.*[5] and Jarvis.[6]

When the propagation medium is homogeneous, the forward differentiated wave is absent,[6, 7] but the near-end crosstalk is of the same form as in the heterogeneous case. A typical homogeneous medium is the multi-layer transmission system, in which the signal conductors are buried in the dielectric between two earth planes. Crosstalk in such systems is analysed by Connolly[9] who indicates the similarity between cross-coupling in a homogeneous medium and the directional electromagnetic coupler.[10] Therefore, papers concerned with the theory of directional couplers may prove to be useful references when studying crosstalk in homogeneous media.

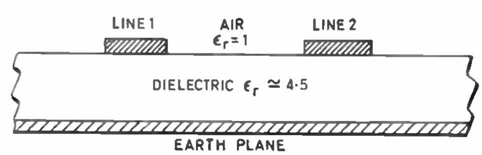


Fig. 2. Typical microstrip transmission system.

There are many situations where lines may be considered to be short, and lumped analysis is furnished by Rössing and Walther[11] utilizing a Fourier integral technique. A mathematical expression is derived for the crosstalk voltage developed at the input of gate 2b on the quiet line. In a worked example, the computed waveform is shown to be in excellent agreement with that observed experimentally. The usefulness of the above techniques depends upon the existence of methods of determining the line parameters, particularly self and mutual inductances, and capacitances.[4, 12-14] The analysis of multiple crosstalk, i.e. crosstalk between several lines, may be quite complex. It may be shown that when several lines crosstalk into a single line, the principle of superposition applies. In the worst case, the crosstalk voltages produced by each source may be directly additive in the quiet line.

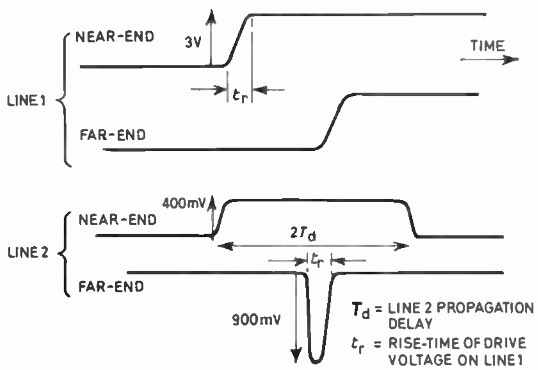


Fig. 3. Typical waveforms for system of Fig. 2, with matched lines.

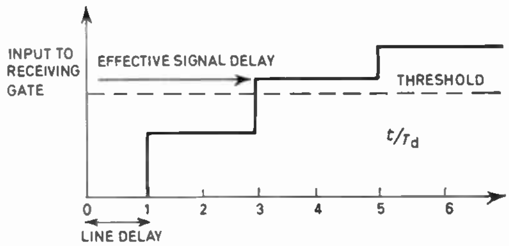


Fig. 4. Increase in signal delay time due to reflections.

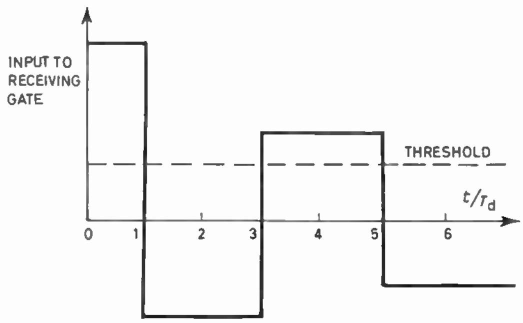


Fig. 5. Oscillations at input of receiving gate.

Crosstalk analysis is further complicated by the fact that lines in digital systems are rarely matched, i.e. they are not terminated in their characteristic impedance. The mismatch is usually quite marked and analysis is made more complex by the non-linear nature of the input characteristics of the commonly used digital circuits. An analysis technique due to Amemiya[15] is useful in problems of multiple crosstalk and mismatch.

In conclusion, it may be said that the far-end crosstalk is particularly important since this is the stimulus seen by the input of the quiet line receiving end gate. Such a stimulus may be responsible for false operation of this gate, especially when high-speed circuits are being used.

3.2. *Reflections*

It is well known that if a transmission line is not terminated in its characteristic impedance, reflections will be produced. Transmission lines in digital systems are rarely matched, the impedance presented to the lines by the input or output of the digital circuits usually being very high or very low.

When considering interconnections in digital systems,[5] a line is deemed to be electrically long if the signal transition time is less than 2 × (line propagation delay). Therefore, if the signal transition time is very short, even physically short lines may produce serious reflections. Jarvis[6] analyses the reflections for all possible combinations of sending-end and receiving-end termination. Two detrimental effects upon signal propagation are shown to occur. Examples of these are given below.

3.2.1. Sending-end gate switches from LOW to HIGH

The threshold of the receiving gate represents the input level above which the output of this gate begins to change. Serious reflections cause the signal delay to exceed the line delay (Fig. 4).

3.2.2. Sending-end gate switches from HIGH to LOW

This oscillatory phenomenon is referred to as line ringing. It can be seen that the form of the oscillations may be such that the input of the receiving gate crosses the threshold illegitimately, e.g. between $3(t/T_d)$ and $5(t/T_d)$ in Fig. 5. Therefore there is the possibility of such oscillations causing a false transition of the receiving gate's output. Hence, oscillations are classified as a form of noise. A detailed analysis of reflections between logic circuits is quite complex due to the non-linear nature of the circuits. A useful technique is the graphical method due to James,[16] which utilizes the input and output characteristics of the particular logic circuits and a 'load line' to represent the characteristic impedance of the interconnecting line.

3.3. *Earth Noise and Supply Line Noise*

Within a digital system all the circuits are linked by the common or earth line. An action in one part of the system may have a detrimental effect on another part of that same system, due to coupling via the earth line. One of the most frequent occurrences is illustrated in Fig. 6, which shows two consecutive gates in the system. When a step of current is switched into the earth line (due to circuits switching elsewhere in the system) a voltage V_e is generated as shown, whose amplitude is dependent upon the resistance and self inductance of that portion of the earth line between the two gates.

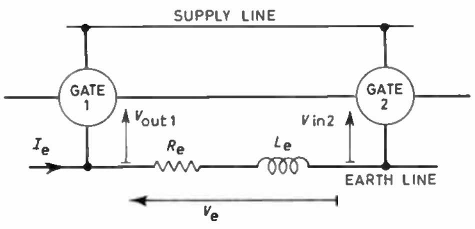


Fig. 6. Cause and effect of earth noise.

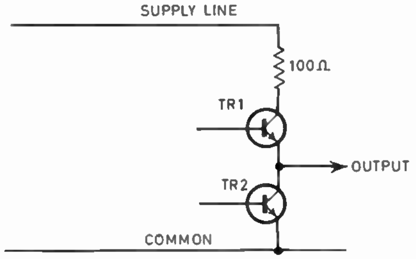


Fig. 7. 'Totem pole' output stage.

The voltage V_e is referred to as the earth noise and, in certain circumstances, is responsible for modifying the signal at gate 2 input. The current drawn by most logic gates depends upon their logical state. Therefore, in a digital system, widely varying amounts of current will be drawn from the supply line. Depending on the supply line characteristics, this may produce fluctuations in supply voltage, which are referred to as supply line noise. Circuits having 'totem pole' output stages, as shown in Fig. 7, may cause serious supply line noise, since charge storage effects can cause both TR1 and TR2 to be 'on' at the same time.

4. D.C. Noise Margins

It was shown in the previous section that system noise invariably takes the form of pulses or spikes. Therefore, in order to determine the noise rejection properties of digital circuits it is necessary to determine their response to such waveforms. Both the d.c. and transient response are of importance, but when d.c. noise margins are under discussion, only the *amplitude* of the noise and the *d.c.* response of the circuits are considered.

4.1. *Signal Propagation*

Complex digital systems are formed by the interconnection of a large number of identical basic circuits. Such circuits perform a simple individual function (e.g. NOR/NAND gate, J-K flip-flop, etc.), whereas the overall system performs a complex function. Within such systems, long series of cascaded gates occur very frequently, therefore the effect of noise upon such an arrangement is of considerable interest. In fact, a quasi-infinite cascade of single input gates is often used when studying the effects of signal noise or when formulating definitions for d.c. noise margins. The justification for the use of such a model is given by Gascoigne.[17]

The propagation of signals through the cascade of gates may be conveniently studied by using the voltage transfer characteristic of each gate concerned. However, each gate may be operating under different conditions of fan-out, temperature and supply voltage, also each gate may possess different characteristics due to production spreads. The effect of these variables can be taken into account by using a toleranced or 'broad' transfer characteristic, which is representative of any gate of the family used, when that gate is operating within the specified operational limits. The toleranced transfer characteristic can often be obtained from manufacturers' data sheets; a typical example is shown in Fig. 8.

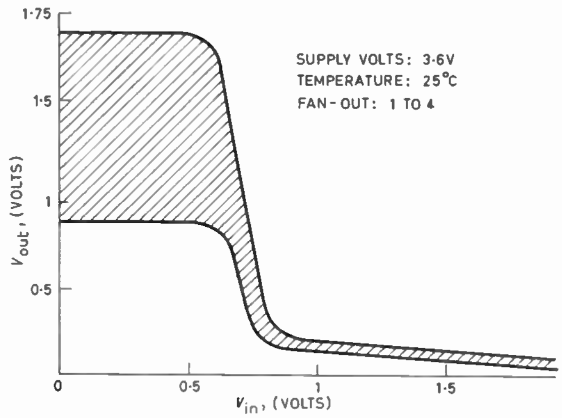


Fig. 8. Toleranced transfer characteristic for a RTL NOR/NAND gate.

The processing of a signal, as it propagates through the cascade of gates, may be illustrated on a diagram which is formed by the combination of two transfer characteristics on common axes.[18] Such a diagram is often known as a logical compatibility diagram and a typical diagram for RTL gates is shown in Fig. 9, together with the cascade of gates.

The areas formed by the overlap of the two characteristics, namely regions I, II and III in Fig. 9, represent conditions of equilibrium within the cascade. It can be shown that,[19] regions I and III represent a stable equilibrium, but that regions II represents an unstable equilibrium. The effect of this as far as signal propagation is concerned is that:

(a) Signals entering the cascade (at gate 1 input), of amplitude less than V_x, are converged to the stable operating region I.

(b) Signals entering the cascade of amplitude greater than V_y are converged to the stable operating region III.

(c) Signals entering the cascade, of amplitude greater than V_x but less than V_y, may be converged to region I or region III or remain in region II, the unstable region.

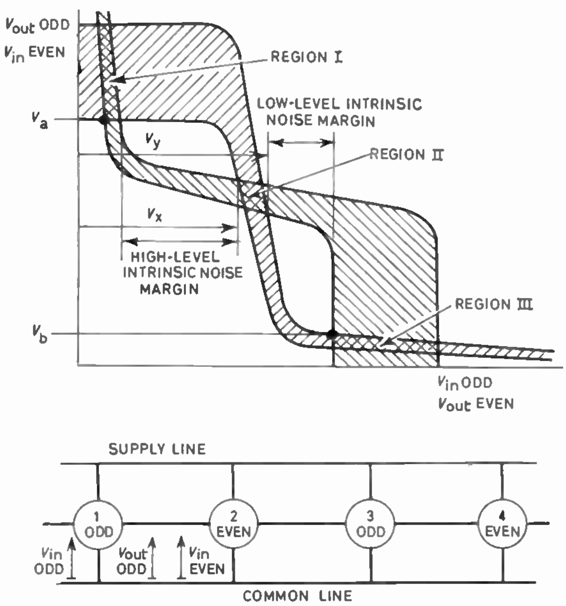


Fig. 9. Logical compatibility diagram and cascade of gates.

Such action is referred to as discriminative logical convergence or signal quantization, and is of importance in the study of d.c. noise margins. The processing of a signal of amplitude less than V_x as it propagates through the cascade is shown in Fig. 10. Discriminative logical convergence occurs in cascades of inverting gates and non-inverting gates, and in practice, a 'worst case' input signal of amplitude V_x or V_y will be converged to region I or region III within six to nine gates from gate 1.

4.2. *Definitions of D.C. Noise Margin*

When d.c. noise margins are discussed, it is always with respect to noise on the signal or input line. It will

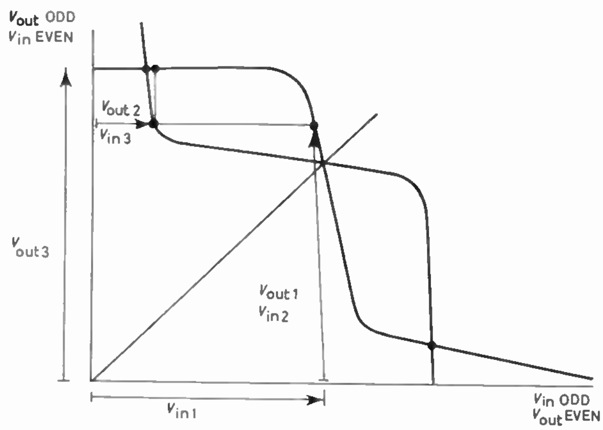


Fig. 10. Processing of a signal through a cascade of gates.

be shown that earth noise and supply line noise can be related to signal noise, therefore no separate definition is needed for noise margins with respect to these forms of noise.

The majority of definitions of d.c. noise margin are based on the voltage transfer characteristic, since this represents the d.c. response of the logic gates under consideration. It can be seen from Fig. 9 that within a long cascade of gates, limits are naturally set on the HIGH and LOW levels. The limits which are of interest in noise margin definitions are the minimum HIGH and the maximum LOW, shown as V_a and V_b respectively in Fig. 9. Having established the limits of the HIGH and LOW output levels, it is necessary to define the limits of the input level, i.e. the input levels beyond which a false transition of a gate's output is deemed to have been caused. The criteria in deciding these limits of input level vary widely from author to author, some with reasoned justification, others with apparently no justification. In the general case, the various input and output voltages can be represented as in Fig. 11.

The points X_H and X_L are known as the transition points. From the diagram, the noise margins may be

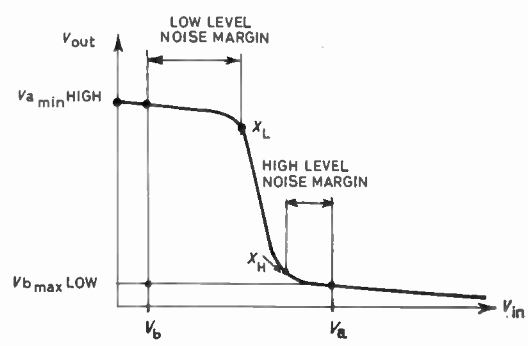


Fig. 11. Nominal transfer characteristic with input and output level limits.

defined for a gate which is in the HIGH or LOW state, where the terms HIGH and LOW refer to the input level.

$$\text{LOW level noise margin} = X_L - V_b$$

$$\text{HIGH level noise margin} = V_a - X_H$$

This may be expressed in either volts or millivolts.

In this definition, noise margin represents the maximum amplitude of noise, which when added to the legitimate input signal (minimum HIGH or maximum LOW), causes the output to reach the appropriate transition point.

It should be noted that the LOW level noise margin is applicable to noise of a positive polarity, while the HIGH level noise margin is applicable to noise of a negative polarity. The noise margins illustrated are only nominal; 'worst case' margins must be based on a toleranced transfer characteristic.

The definitions of noise margin due to Tempel[20] and Luecke[21] follow the method outlined above. Tempel considers the points of unity slope on the transfer characteristic to be indicative of the transition points, whereas Luecke's criteria are not outlined. Texas Instruments' method[22] of defining noise margins for Series 74 TTL is also similar to that above. Under 'worst case' conditions it is stated that:

0·8 V input guarantees a 2·4 V minimum output

2·0 V input guarantees a 0·4 V maximum output.

Therefore

HIGH level noise margin = (2·4 − 2·0) = 400 mV

and

LOW level noise margin = (0·8 − 0·4) = 400 mV.

A method due to Rhoades[23] advocates plotting the transfer characteristic across a *pair* of gates. It is shown that this technique highlights transition points, which are not apparent from the transfer characteristic of a single gate. Noise margins are then defined in a manner similar to the previous examples. However, this method is based on the assumption that no noise is injected into the interconnection between the two gates. This assumption limits the usefulness of the method and if it is carried to its obvious conclusion, i.e. plotting the transfer characteristic across a quasi-infinite cascade of gates, the definition yielded is similar to the *intrinsic* noise margin, as defined by both Gascoigne[17] and Hill.[19]

Other commonly used definitions of noise margin are those based on circuit conditions. These methods are particularly useful for current-sinking logic and an example of the application of such a method to DTL is given by Ricks.[25] This particular method is based on the criterion that the output transistor of the gate must never be driven into the active region, where it may respond to any noise appearing at the input. Therefore, under static conditions, the output transistor must be either in the cut-off region or the saturation region.

All the methods considered so far have apparently been concerned with defining the noise margins of a gate in isolation. However, the ultimate aim must be the definition of noise margins for a gate when it is part of a digital system. Such considerations are noted by Gascoigne[17] and Hill,[19] who consider the effect of noise on the system as a whole. This is not to say that such considerations are not implicit in the definitions used by other authors but, as such, are not mentioned in their papers.

Both Gascoigne and Hill show that if noise occurs at the input of only one gate within the cascade, then the effective noise margins are given by the margins between the output level limits (min. HIGH and max. LOW), and the edges of the unstable region II (Fig. 9). This is so, since noise of sufficient amplitude to cause the total input signal to cross the region II, will cause a false state to propagate through the cascade, due to the action of discriminative logical convergence. Such noise margins are referred to as *intrinsic* and may be said to have limited applications.

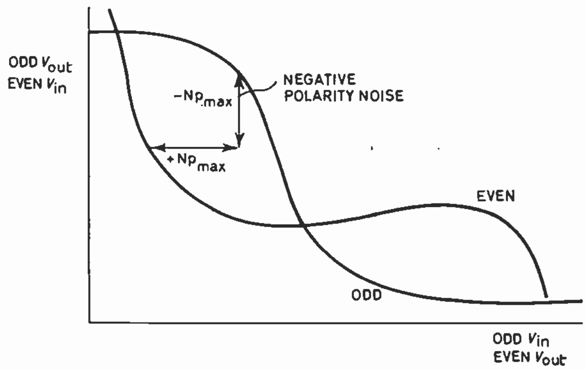


Fig. 12. Definition of d.c. noise margin based on logical compatibility diagram.

The most useful definition of noise margin is that which makes allowance for equal amplitudes of noise occurring at the input of *each and every gate* within the cascade. Such a definition is said to yield the 'worst case' noise margins, and in a cascade of inverting gates it allows for the noise at the inputs to the odd gates being of the opposite polarity to that at the inputs to the even gates. The definition due to Gascoigne is based on the transfer characteristic of non-inverting gates and extended to include inverting gates, whereas that due to Hill is directly applicable to either type of gate and is based on a logical compatibility diagram, as shown in Fig. 12. Hill comments that the majority of noise margin definitions in use at present yield worst-case values which are safe but pessimistic, due to the form of the transfer characteristic of the commonly used logic circuits. The method advocated by Hill yields the optimum value of 'worst case' noise margin irrespective of the form of the circuits' transfer characteristic.

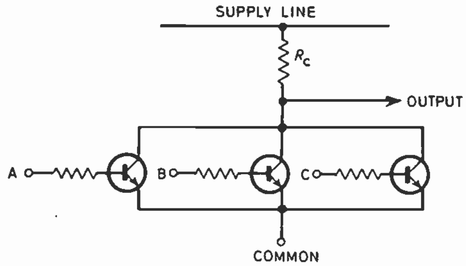
4.3. *Margins with Respect to Earth Noise and Supply Line Noise*

It has been shown that both earth and supply line noise may occur in systems, thus it is necessary to be able to ascertain the noise margins of circuits with respect to these forms of noise. For most circuit configurations, the margins with respect to earth and supply line noise may be related to the signal noise margins. As an example, consider the RTL gate shown in Fig. 13(a).

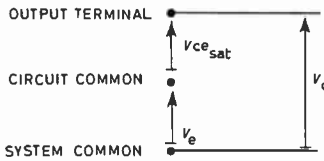
Firstly, consider the earth noise.

(a) With all transistors 'off', i.e. HIGH output level, the output is not affected by noise at the circuit's common terminal. The output voltage is a function of the supply line voltage, the value of the common collector resistor and the output current which flows in an external load.

(b) With one or more transistors saturated, i.e. LOW output level, the output is effectively clamped at $V_{ce(sat)}$ above the potential of the circuit's common terminal. Any variation in potential of the circuit's common terminal, due to earth noise, causes the same variation to appear at the output terminal, as shown in Fig. 13(b).



(a) RTL gate.



(b) Effect of earth noise on gate output.

Fig. 13.

Thus it may be seen that noise appearing at the common or earth terminal of one gate will appear as signal line noise to the following gate. Therefore, the margin with respect to earth noise is identical to the LOW level signal noise margin.

Considering supply line noise;

(a) With one or more transistors saturated, i.e. LOW output level, the output is unaffected by supply line variations.

(b) With all transistors 'off', i.e. HIGH output level, the output will follow supply line voltage variations if a constant current is drawn from the output terminal. Under such conditions, the noise margin with respect to supply line noise is identical to the HIGH level signal noise margin.

4.4. *A.C. or Pulse Noise Susceptibility*

It has been shown that the noise generated in systems takes the form of pulses or spikes, therefore to make a true assessment of the noise rejection capability of digital circuits their transient response to time varying noise must be considered. The analysis of complete circuits can be very complex, and it is understandable that few authors present methods of assessing pulse noise susceptibility. The main problem seems to be in finding a convenient method of presenting such information. The problem of analysing the transient response of integrated circuits is complicated by the existence of parasitic elements (mainly transistors) within the silicon chip.

Yao[24] employs a method based on charge-control parameters and illustrates his method with an example involving a discrete component DTL gate. Such a method is limited to simple circuit configurations such as RTL and DTL, but should prove useful for these types of circuit.

Luecke[21] advocates an experimental method for determining the pulse noise susceptibility, using a defined pulse as a stimulus and observing the output response. Irrespective of the method of specifying the pulse noise susceptibility, it may be said that as the noise pulse width decreases, the amplitude of noise that can be tolerated increases.

5. Acknowledgments

This paper is based on material collected during a research programme at Letchworth College of Technology, Hertfordshire. The author is indebted to the Principal of the College for the research facilities afforded to him. Sincere thanks are also due to Mr. K. J. Dean, Head of the Department of Science and Electrical Engineering for his helpful suggestions and criticisms during the preparation of this paper.

6. References

1. Bloch, E., 'The engineering design of the Stretch computer', Proc. Eastern Joint Computer Conference, pp. 48–58, 1959.
2. Prywes, N. S., Lukoff, H. and Schwarz, J., 'UNIVAC–LARC high speed circuitry; case history in circuit optimisation', *Trans. Inst. Radio Engrs on Electronic Computers*, **EC-10**, pp. 426–38, September 1961.
3. Poppelbaum, W. J. and Wiseman, N. E., 'Circuit design for the new Illinois computer', Digital Computer Laboratory, University of Illinois, Report No. 90, August 1959.
4. Gray, H. J., 'Digital Computer Engineering', Chapter 9 (Prentice-Hall, Englewood Cliffs, 1963).
5. Feller, A., Kaupp, H. R. and Di Giacomo, J. J., 'Crosstalk and reflections in high speed digital systems', Proc. AFIPS Fall Joint Computer Conference, pp. 511–25, 1965.
6. Jarvis, D. B., 'The effects of interconnections on high speed logic circuits', *Trans. Inst. Elect. Electronics Engrs*, **EC-12**, pp. 476–88, October 1963.
7. Motorola Integrated Circuits Design Course Notes, Section 26c, 1963–64.
8. Garth, E. C. and Catt, I., 'Ultra high speed integrated circuit interconnections', *Design Electronics*, **5**, No. 1, p. 46, October 1967.
9. Connolly, J. B., 'Cross coupling in high speed digital systems', *Trans. I.E.E.E.*, **EC-15**, No. 3, pp. 323–7, June 1966.
10. Oliver, B. M., 'Directional electromagnetic couplers', *Proc. I.R.E.*, **42**, pp. 1686–92, November 1954.
11. Rössing, J. F. and Walther, J. E., 'Calculation of cross-coupled noise in digital systems', *Trans. I.E.E.E.*, **EC-16**, No. 1, pp. 14–7, February 1967.
12. Grover, F. W., 'Inductance Calculations, Working Formulas and Tables' (Dover, New York, 1965).
13. Kaupp, H. R. 'Characteristics of microstrip transmission lines', *Trans. I.E.E.E.*, **EC-16**, No. 2, pp. 185–93, April 1967.
14. Schwarzmann, A., 'Microstrip plus equations adds up to fast designs', *Electronics*, **40**, No. 20, pp. 109–12, 2nd October 1967.
15. Amemiya, H. 'Time domain analysis of multiple parallel transmission lines by means of equivalent circuits', *Electronics Letters*, **3**, No. 1, pp. 13–5, January 1967.
16. James, J. B., 'Problems of interconnecting high-speed integrated circuits', Proc. 'Microelectronics in Equipment' Conference, London, November 1966, (Paper No. B-6).
17. Gascoigne, R. M., 'Noise immunity of logic circuits', *ibid.*, (Paper No. C-2).
18. Lo, A. W., 'A comprehensive view of digital integrated electronic circuits', *Proc. I.E.E.E.*, **52**, No. 12, pp. 1546–50, December 1964.
19. Hill, C. F., 'Definitions of noise margin in logic systems', *Mullard Technical Communications*, **9**, No. 89, pp. 239–45, September 1967.
20. Tempel, J. A., 'Noise immunity in digital solid integrated circuits', *Design Electronics*, **3**, p. 34, June 1966.
21. Luecke, G., 'Noise margins in digital integrated circuits', *Proc. I.E.E.E.*, **52**, No. 12, pp. 1565–71, December 1964.
22. Texas Instruments, Series 74 Application Report.
23. Rhoades, W. T., 'New test techniques for digital integrated circuits', Proc. National Electronics Conference, pp. 215–9, 1966.
24. Yao, F. C., 'Interconnection and noise immunity of circuitry in digital computers', *Trans. I.E.E.E.*, **EC-14**, No. 6, pp. 875–80, December 1965.
25. Ricks, R., 'Noise immunity; what it really means', *E.E.E.—Circuit Design Engineering*, **14**, pp. 90–1, April 1966.

7. Appendix

Table 1

Data on commercially available digital integrated circuits
(All parameters quoted are for similar gate circuits and represent typical values at 25°C.)

Manufacturer	Supply V_{cc} (volts)	Logic swing (volts)	Fan-out	Propagation delay (ns)	Power dissipation (mW)	Family
Motorola	3·0	1·0	4	24	24	RTL
Fairchild	3·6	1·1	5	12	12	RTL
Mullard	6·0	5·2	8	31	11	DTL
R.C.A.	4·0	3·3	6	60	2·3	DTL
Texas	5·0	2·0	10	13	10	TTL
Mullard	5·0	2·0	10	13	10	TTL
Motorola	5·2†	0·8	25	5	47	ECL

† Additional bias voltage of 1·15 V is usually derived from bias driver module.

Manuscript first received by the Institution on 7th February 1968 and in final form on 13th August 1968. (*Paper No. 1218/CC25.*)

U.D.C. 546.96: 621.316.849

Ruthenium Resistor Glazes for Thick Film Circuits

By

G. S. ILES†

Reprinted from the Proceedings of the Joint I.E.R.E.–I.E.E.–I.S.H.M. Conference on 'Thick Film Technology', held at Imperial College, London, on 8th to 9th April 1968.

Summary: The rapid development of thick film micro-circuits has created a need for preparations that will provide resistor films on a variety of substrates. Until recently the majority of the resistive preparations available required a firing profile with a peak temperature around 750°C, necessary to complete complex chemical changes within the glaze resistor. This not only imposed the need for very close control of the firing cycle and furnace atmosphere to ensure reproducible values, but limited the choice of substrate to materials, such as 95–98% alumina ceramics, capable of withstanding this temperature. A prerequisite to attaining consistent values is the surface condition of the substrate. Relatively low-cost materials like mica and glass, which inherently possess flat, highly finished surfaces, are excluded, since they either exfoliate (dehydrate) or deform at temperatures exceeding 650°C.

To overcome these limitations and take advantage of the mechanism of conduction through ruthenium dioxide a series of glaze resistor preparations has been developed. The objective was an ink based on a chemically-inert system that would be far less dependent on firing conditions than those hitherto available. To be viable the system has to be capable of producing a wide range of resistivities; possess acceptably low temperature coefficients; contain the spread of resistance values within narrow limits and be economically competitive. The use of valency control and the techniques adopted to achieve these requirements are described.

1. Introduction

Recent years have witnessed a mounting interest in thick film microcircuits, stimulated by the relative ease and low cost with which they can be produced.

There is now little doubt that within the next decade the proportion of electronic equipment featuring thick film circuits in their construction will substantially rise and probably exceed that manufactured from discrete components.

Conventional circuits containing discrete components are normally assembled on a printed circuit board. In thick film technology however, the circuit elements are applied as inks and subsequently fired on to the substrates, several of which may be co-assembled to form a module.

While the circuits so produced are bulkier than their thin film counterparts, they take advantage of the simple, well-established screen printing skills offering versatility and lower manufacturing costs.

Of the circuit components which can be fabricated by this method, the greatest demand is for resistive elements. This dictate has been resolved by the development of resistor inks usually comprising suspensions of a powdered glaze (frit) and one or more noble metal powders in an organic vehicle. After application to the substrate, the print is dried and fired to burn away the organic material, fuse the glaze component and complete any other reactions necessary.

Various sheet resistivities are made available by varying the composition of the ink. Close limits of resistance are achieved by trimming the fired film, generally by abrasion with a carefully monitored alumina blast.

Until recently the majority of the resistive preparations available required a firing profile with a peak temperature in excess of 700°C, necessary to complete complex chemical changes in the metallic component of the resistor glaze. This not only imposes the need for very close control of the firing cycle and furnace atmosphere to ensure reproducible values, but limits the choice of substrate to materials, such as 95–98% alumina ceramics, capable of withstanding this temperature. The surface condition of the substrate is a prerequisite to attaining consistent values. Relatively

† Johnson, Matthey & Co. Limited., Research Laboratories, Wembley, Middlesex.

low-cost materials like mica and glass, which inherently possess flat, highly finished surfaces, are excluded, since they either exfoliate (dehydrate) or deform at these temperatures.

To overcome these limitations a series of glaze resistor preparations containing ruthenium dioxide was developed. The objective was an ink based on a chemically inert system that would depend on the firing stage to fulfil two simple functions, namely, burning out the organic matter in the vehicle and sintering the metal oxide/glaze particles. By carrying out these operations at lower temperatures in the region of 600°C, advantage may be taken of the less expensive, high surface quality materials.

To be viable the resistor material had to satisfy four other conditions: (1) The system had to be capable of producing a wide range of stable resistors. (2) The films had to possess acceptably low temperature coefficients of resistance. (3) The ruthenium component had to be used as economically as possible. (4) The spread of resistance values obtained from any nominal value had to be as small as possible.

2. Resistor Material

2.1 *Conduction through Ruthenium Dioxide*

Ruthenium dioxide is a blue-black, electrically conductive crystalline solid possessing the rutile structure. Unlike palladium oxide, it can be heated in air to about 1000°C without undergoing any chemical change, and is insoluble in a wide variety of frit and glass compositions.

So-called ruthenium dioxide is seldom, if ever, obtained absolutely pure or as stoichiometric RuO_2. It is usually partially defective in oxygen, probably with a corresponding amount of Ru^{3+} in place of Ru^{4+} in the crystal lattice.[1] These deviating valencies suggested a mode of electrical conduction within the oxide, but since the composition of the oxide showed variations from batch to batch, the maintenance of reproducible resistance values in the glaze film might be expected to pose a problem. It was soon demonstrated that valency control within narrow limits was necessary if repetitive, stable resistors based on ruthenium dioxide were to be developed.

2.2 *Control of Valency*

2.2.1 In semiconducting oxides

Work on the control of deviating valencies in semiconducting oxides, in particular nickel oxide, was reported by Verwey[2] and his co-workers. It was shown that introduction of suitable ions into the lattice structure of a variable oxide could, without deforming it, balance the ions of deviating valency already within the lattice and still maintain overall neutrality. For example, Verwey obtained a composition $Li_\delta^+ Ni^{2+}_{(1-2\delta)} Ni^{3+}_\delta O$ by calcining lithium carbonate with nickel oxide, at 1200°C under oxidizing conditions. The product had the same structure as nickel oxide, but with a smaller unit cell, and the Ni^{3+} content was broadly equivalent to the amount of lithium oxide added.

This suggested that valency variations in ruthenium dioxide might be controlled by a similar 'doping' technique, leading to a better reproducibility from batch to batch, together with a measure of control over both resistivity and temperature coefficient.

2.2.2 In ruthenium dioxide

In order to make use of this technique it was necessary from theoretical considerations to introduce a pentavalent ion into the lattice structure to balance the already established Ru^{3+} ions and maintain overall neutrality. Niobium pentoxide was found to be a suitable oxide with which to dope ruthenium dioxide and various amounts were successfully introduced into the lattice. The resultant oxide system was shown to obey Vegard's Law (Fig. 1), which states in effect that the extent of the change in lattice parameter of the host oxide is proportional to the molecular percentage of added dopant. This linearity provided the first stage in process control by offering a useful means of monitoring the composition by x-ray diffraction prior to the manufacture into a resistor ink.

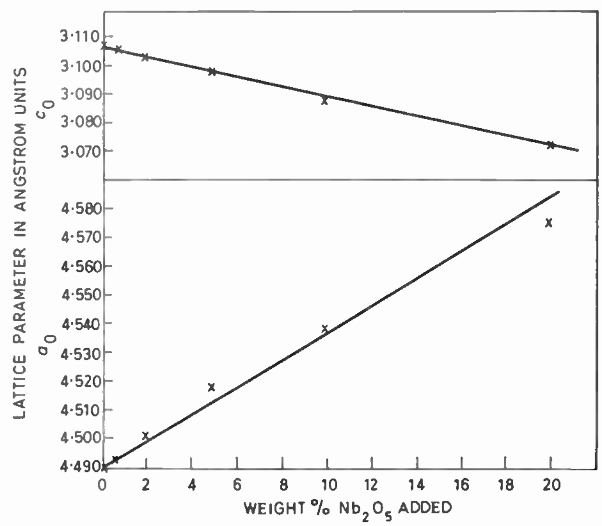


Fig. 1. Effect of additions of Nb_2O_5 on the lattice parameters of RuO_2 showing conformity to Vegard's law.

Moreover, since the temperature coefficient of resistance of ruthenium dioxide is metallic in nature and strongly positive, introduction of a compensating oxide might be expected to exert a negative influence on the temperature coefficient. Thus control of

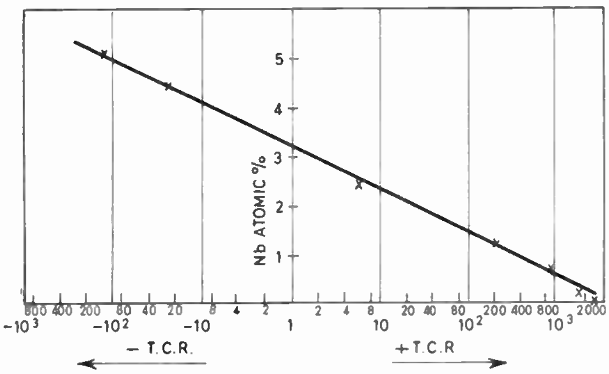


Fig. 2. Dependence of temperature coefficient of resistance (t.c.r.) on dopant level in host oxide.

temperature coefficient of resistance in addition to resistivity would be achieved. Experimental evidence confirmed the correctness of this assumption and indicated that a measure of control could be exercised over the temperature coefficient by varying the level of the dopant oxide. It was shown that the resistivity of the film was essentially governed by the ratio of doped ruthenium dioxide to glass, and the temperature coefficient by this same ratio in conjunction with the atomic percentage of niobium contained in the host lattice. For example, ruthenium dioxide glaze films in a wide range of resistance values were found to have positive coefficients in excess of 1000×10^{-6}/degC. As the atomic percentage of niobium in the calcine was increased the temperature coefficient decreased, reaching a negative value of 100×10^{-6}/degC at a niobium content of 5% atomic (Fig. 2).

Since the mixed oxides are reacted by calcination before incorporation in the resistor preparation, little or no reaction should occur when the preparation is subsequently fired on the substrate. Measurements of the sheet resistance and temperature coefficient of resistance showed that these properties were not unduly affected by variation in the time of firing or in the temperature and atmosphere in the furnace. For example, neither a tenfold variation in time for a peak temperature of 650°C, nor a temperature fluctuation of ±30degC produced any significant effect on either parameter. A typical firing profile is shown in Fig. 3.

Silver powder was found to be a useful addition to ruthenium dioxide based preparations. Up to 60% of the ruthenium dioxide could be replaced by silver without adversely affecting the temperature coefficient provided a balance was struck between the niobium and silver contents. The positive coefficient of the latter compensated the negative influence of the niobium pentoxide. Thus silver provided an additional means of controlling the temperature coefficient of resistance in addition to reducing the cost of the resistor preparation.

It is possible however, that the presence of silver in the resistor film makes the system more sensitive to time variations during passage through the peak region of the firing profile. Further work is proceeding to establish the limits of the various parameters.

2.3 *The Glaze Component*

During the evaluation of the effect of various dopants on the electrical properties of ruthenium dioxide, it was established that the glaze component also produced a significant effect, particularly on the temperature coefficient of resistance of the fired film. For example, melts of differing compositions from the family of lead borosilicate glasses, whether simple or complex, promoted high positive temperature coefficients. These varied from $500–2000 \times 10^{-6}$/degC depending upon whether doped or undoped ruthenium dioxide was used in the resistor preparation.

Better results, with $\pm 250 \times 10^{-6}$/degC, were obtained from lead free melts of the zinc and cadmium borosilicate type.

It was also shown that glasses consisting essentially of bismuth borate were very prone to reduction during the preparative stage, and that resistor films subsequently made from these melts also possessed highly positive temperature coefficients.

These facts suggested that glasses which were subject to change by reduction were undesirable in glaze resistors.

Apart from chemical consideration of the melting process, it is considered to be good practice in the glass industry to allow sufficient time for the melt to out-gas and homogenize before pouring commences. This practice has been adopted throughout the investigation of ruthenium dioxide resistor glazes, and

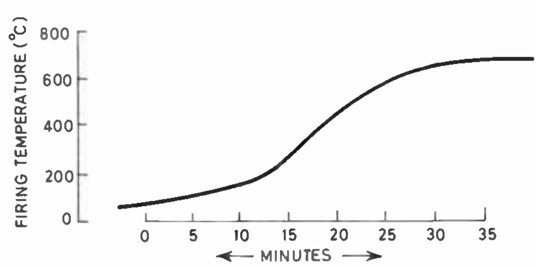


Fig. 3. Typical firing profile for ruthenium resistor glazes.

Notes: (1) Total throughput time = 72–90 min.
(2) Dwell in 600–650°C range = 20–22 min.
(3) Maximum gradient should not exceed 45degC/min in heating cycle.
(4) Cooling profile should simulate heating profile as closely as possible.

a high degree of consistency has been obtained from batch to batch of the glaze component. It has therefore become part of the manufacturing process.

3. Electrical Properties

At present ruthenium oxide preparations are available commercially[3] covering the range from 5 to 100 000 ohm/sq/mil firing at 600°C upwards.

Originally, it was intended to market seven standard preparations covering the full range, and obtain intermediate values by blending the compositions nearest to the desired resistance.

Experience has shown that the operation of blending high viscosity materials (e.g. screening inks) can only be satisfactorily performed with the appropriate equipment especially where the various grades are similarly dark in colour and thereby eliminate any visual aid to homogeneity.

Since the extent of the spread of resistance must primarily depend upon the distribution of the various components in the ink, it would be unsatisfactory to condemn an otherwise suitable material purely because the blending operation had been carried out inefficiently. Therefore, it is recommended that inks should be supplied to the values suited to user's requirements. In this way, a prime variable would remain within the control of the resistor ink manufacturer.

The use of precise printing equipment coupled with suitable skill, a controlled firing profile and a carefully formulated resistor ink should maintain values to within $\pm 10\%$ of nominal. Experience in handling the various equipments and materials should eventually allow much closer tolerances to be maintained. For example, where positive aspect ratios are used (i.e. the length exceeds the width of the print) $\pm 5\%$ or less should be attainable, whereas with negative ratios (i.e. the width exceeds the length) the precision of the geometry will probably outweigh the other parameters and maintenance of low tolerances will become more exacting.

It is important to remember that there is a threshold value for the fired film thickness below which anomalous values of resistance may occur. In the case of ruthenium oxide resistor glazes fired films should not be allowed to fall below 12·5 μm (0·0005 in) thickness because some areas of the print may approach uniparticulate thickness, giving rise to abnormally high resistance spots, and an associated degradation of the temperature coefficient.

Temperature coefficients in the range $\pm 100 \times 10^{-6}$ per deg C can be expected with sheet resistivities from 50 – 5000 ohm/sq/mil, as a general rule, with an improvement to within $\pm 50 \times 10^{-6}$/degC where particular attention is paid to the geometric and deposition precision of the print, followed by a firing schedule which is controlled and reproducible.

At the higher end of the resistance range, i.e. 5000–100 000 ohm/sq/mil, temperature coefficients from -50×10^{-6}/degC to -200×10^{-6}/degC may be expected, subject to the observance of precautions already outlined.

At the lower end of the scale, i.e. 5 ohm/sq/mil, temperature coefficients up to $+250 \times 10^{-6}$/degC are normal, but with further development it is expected that a sheet resistivity of 1 ohm/sq/mil will not exceed $+250 \times 10^{-6}$/degC, which may be further reduced to $+200 \times 10^{-6}$/degC or less.

The temperature coefficients associated with these various sheet resistivities may satisfy most applications, but may not offer the optimum performance for certain specific applications. For example, there may be a case for inks of a particular sheet resistivity possessing negative or positive temperature coefficients to fulfil a particular circuit design. The technique adopted in this family of ruthenium resistor glazes allows adjustments of this nature to be met. The exact limits of sheet resistivity for which compensated temperature coefficients are available have yet to be defined, but requests can be considered on their own particular merits. The following example illustrates the concept. An ink with a sheet resistivity of 1000 ohm/sq/mil could have a temperature coefficient ranging from -250×10^{-6}/degC to $+100 \times {}^{-6}$/degC, depending upon the level of dopant in the ruthenium oxide.

Data on the noise level of resistors made from ruthenium oxide inks are still not complete, but indications are that the noise values are independent of the conductor material used as terminal lands. Resistors having lands either prefired or co-fired in silver, gold or gold/ruthenium oxide exhibited noise values of -12dB/decade and $+2$dB/decade for resistance values of 8000 and 29 000 ohms respectively, measured on a QuanTech noise meter.

Preliminary stability measurements show a drift of $<0{\cdot}5\%$ when subjected to a load of 5W/in²/mil at 70°C for 1000 hours, and $<0{\cdot}25\%$ after 2000 hours at 125°C no load. After ten thermal cycles from $+20$°C to $+105$°C drift was $<0{\cdot}1\%$.

Short duration power dissipation tests showed no measurable drift at loadings up to 100W/in²/mil. This value, however, does not represent the limit, since substrate failure occurred at loadings of 130W/in²/mil causing mechanical failure of the resistor film.

4. Rheological Properties

The organic medium in which the solid components are dispersed was formulated to give a paste with rheological properties suitable for screen printing.

The solid components are processed through a series of carefully controlled milling stages. This allows a high degree of dispersion to be maintained to a consistent standard, thus avoiding over-milling and its consequent degrading effect on the electrical properties.

The resultant ink possesses sufficient body to prevent separation of the various components during storage, at the same time allowing mesh marks to flow out, while maintaining a sharp profile.

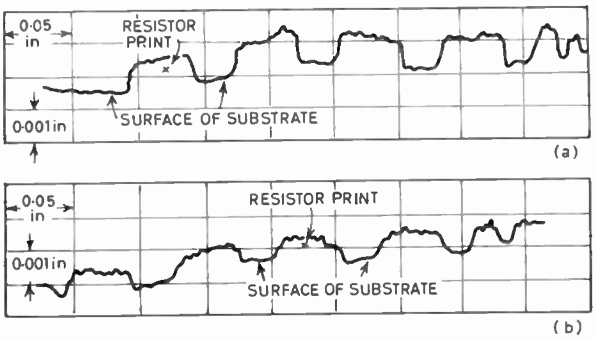


Fig. 4. (a) Profiles of dried but unfired prints from a ruthenium oxide ink.
(b) Profiles of fired prints from a ruthenium oxide ink.

Traces of the topography of both dried and unfired, and fired films using a Talysurf 4 Meter conveniently confirmed these facts (Fig. 4).

The behaviour of an ink during and immediately after the screen printing process profoundly influences the reproducibility and spread of resistance values obtained.

During the screening operation work is done on the ink due to the shearing action of the squeegee passing over the screen mesh, and the viscosity decreases. The extent of this decrease depends upon the work history of the ink, the degree of thixotropy (or more correctly 'false body'), and the recovery time of the ink (that is the time taken for the ink to regain its original viscosity after the shearing stress has been removed).

4.1 *Method of Examination*

These parameters were examined on a Ferranti-Shirley cone-plate viscometer. On this instrument a sample of ink, maintained at a constant temperature (25°C), was subjected to a steadily increasing rate of shear to a predetermined level, then immediately subjected to a reversal of the process—a steadily decreasing rate of shear back to zero.

A graphic trace of the variation in viscosity against shear rate was obtained using an X-Y recorder.

By varying the size of the cone used to apply the stress, the sweep time to maximum shear rate, the shear rate and the range selector (1 to 5), various traces were obtained from which deductions are made on the rheological character of ink.

Typical rheograms from two different types of resistor ink are shown in Fig. 5.

The linear separation of the loop boundaries in the X-direction indicates a measure of the structure of the ink—the greater the separation the greater the structure, which may not be recoverable.

Examination of the traces shows that the recovery time for ink 'A' is very short whereas that for ink 'B' is indeterminate. Thus, in practical terms, during continuous use where the squeegee may perform, say, one printing stroke per second, ink 'B' will continue to change its character and apparently become thinner.

Initially, the prints will be well-defined and show the mesh pattern, which may be retained in the dried film due to poor flow experienced at this stage. As the structure is broken down by the shearing action of the squeegee, well-defined prints with reasonable flow-out will appear. This state will be of a short duration, since the viscosity will continue to decrease from lack of structure in the ink. The resultant prints will lose

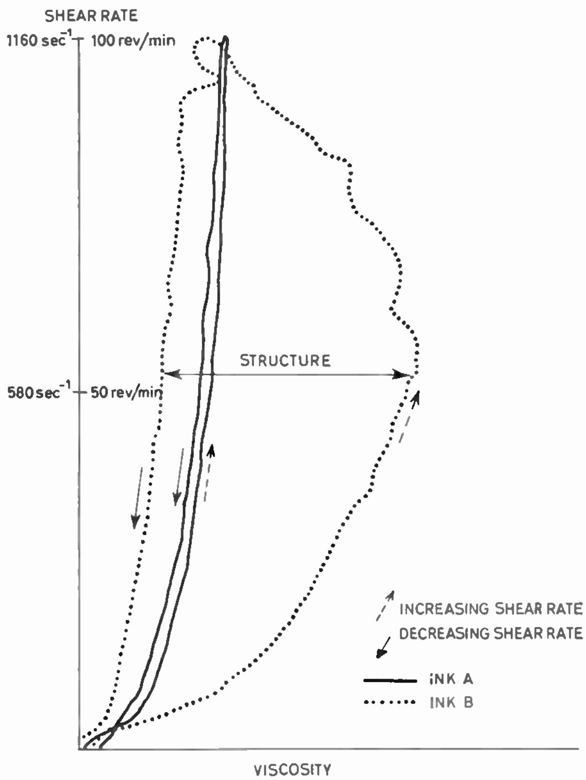


Fig. 5. Rheograms of dissimilar inks.

definition by excessive flow, bubbling in the print is likely to occur, and thickness variations will result in loss of control over resistance values. Thus the maintenance of low-resistance spreads, during a batch run, to $\pm 10\%$ will be difficult if not impossible to achieve. Immediately fresh ink is added to the screeen (topping up) the situation becomes further aggravated, due to the presence of two (or more) viscosity levels in the fluid. This worsening state continues until the prints deteriorate to such an extent that printing has to stop. The screen is cleared of ink and fresh material added, only to repeat the unsatisfactory cycle. The use of ink 'A' will result in greater ease in maintaining tolerances, and increased production.

Ruthenium oxide glaze resistor inks are typified rheologically by the representative curve of ink 'A' shown in Fig. 5.

Other factors can also influence the rheology of ink whilst in the screen, e.g. solvent loss, temperature fluctuations and humidity, to name only three. However, these external influences are of minor importance compared to the factors already discussed.

4.2 *Application to the Substrate*

Deposition of even, well-defined films ranging in thickness after being fired from 50 µm to 12·5 µm (0·002 in to 0·0005 in) may be obtained from screen mesh sizes in the range 100 to 200-mesh. The exact film thickness per mesh size is dependent upon many factors, too numerous to discuss here. The test data already discussed were obtained from a standardized fired film thickness averaging 12·5 µm (0·0005 in) using a 200-mesh stainless steel screen.

The surface of the substrate should be as clean as possible, certainly free from moisture, perspiration, grease, and dust which may adversely affect the adhesion of the circuit element to the substrate.

The surface finish of the substrate is a subject worthy of detailed examination in its own right, and there is as yet very little evidence to suggest the degree of smoothness that is necessary. The final choice would appear to be one of personal preference.

5. Conclusion

It has been shown that the adoption of a doping technique, whereby the lattice structure of ruthenium dioxide is modified, provides a method of controlling the sheet resistivity and temperature coefficient of resistance over a wide range of values.

The inclusion of an homogeneous glass phase which is stable during the melting and subsequent resistor firing stages, enables consistent low temperature coefficients of resistance to be maintained.

Ruthenium resistor glazes prepared and used in the manner described can produce films possessing good electrical properties, whilst at the same time utilizing ruthenium dioxide at the most economic level.

6. References

1. Cotton, F. A. and Wilkinson, G. 'Advanced Inorganic Chemistry', p. 814 (Interscience Publishers, Chichester, Sussex, 1962).
2. Verwey, E. J. W., Haaijman, P. W., Romejin, F. C. and van Oosterhout, G. W. 'Controlled-valency semiconductors', *Philips Research Reports*, 5, pp. 173–87, 1950.
3. Iles, G. S. and Collier, O. N., Johnson, Matthey & Co. Limited, British Patent Application 46910/66.

Manuscript received by the Institution on 31st January 1968. (Paper No. 1219)

U.D.C. 621.372.57

An R-C Synthesis Procedure for *n*th-order Rational Transfer Functions with Maximum Gain Constant and Low Number of Components

By

A. G. J. HOLT, Ph.D., C.Eng., M.I.E.E., M.I.E.R.E.†

AND

K. M. REINECK, Dip.El.Ing., Ph.D., V.D.E.‡

Summary: A passive R-C synthesis for *n*th order rational transfer functions is presented. The synthesis is very simple in application and the networks realized are of the parallel ladder type. The main features of this procedure are maximum gain, minimum number of components, and reasonable component value spreads.

1. Introduction

The far-reaching influence of integrated circuitry in electronic technology has resulted in the revival of interest in R-C circuits. These are now capable of performing many of the digital and linear functions previously performed by conventional circuits using discrete components. The inductor and the ideal transformer have always proved to be unsatisfactory elements as far as their approximation is concerned. No inductor can be regarded as completely lossless and without stray capacitance, and it has not yet been possible to produce them with practical inductance values in thin-film form. The ideal transformer is even less practical than the inductor. Although the characteristics of the ideal transformer have been approximated, they are theoretically unattainable in a passive device.

In avoiding the use of the inductor and ideal transformer one must make use of synthesis methods requiring only resistors and capacitors. If, therefore, networks of the three-terminal kind are considered, having a common ground connection, they can be connected in parallel without an ideal transformer. For a simple R-C ladder network all zeros of transmission are required to be located on the negative real axis of the p-plane. There are, however, applications which require zeros of transmission on the imaginary axis or in the complex frequency plane. This is an impossible task for passive R-C ladder networks having a single transmission path. However, if the possibility of producing zeros of transmission by multiple transmission paths between input and output terminals is considered, the R-C parallel ladder method as suggested by Guillemin[1,2] applies. Guillemin showed that this type of structure is capable of producing imaginary and complex zeros of transmission. Thus it is possible to use R-C networks for the realization of imaginary zeros of transmission, which are usually associated with L-C networks, suggesting filter or certain phase correction applications in the case of complex zeros.

Apart from the Guillemin procedure there exists the Dasher[3] synthesis procedure which is essentially a cascade connection of low-order zero sections, each realizing a pair of imaginary or complex zeros. This procedure may be very time-consuming in its application and little control is possible over the component spread and gain constant. The gain is usually greater than that obtained with the parallel ladder synthesis due to Guillemin. But even so, maximum gain usually cannot be expected. In a recent publication[4] a method for reducing the number of components and obtaining maximum gain in a cascade connection network is given. Also, the component value spread can be controlled to a certain extent.

Guillemin's parallel ladder method is simple in its application to synthesis, but it produces networks having low gain and a large number of components. In the following it will be shown how it is possible to obtain networks having maximum gain and a low number of components.

2. Maximum Gain, a Valid Criterion

From the synthesis of passive ladders, one finds that there are many ways to synthesize and derive networks of different configurations. The factors which influence the structure of the networks are (i) the choice of the divisor polynomial, i.e. the order; (ii) the sequence in which the zeros of transmission are realized; (iii) whether zeros are realized in shunt or series arms and (iv) how many partial pole removal operations are

† Department of Electrical Engineering, The University of Newcastle upon Tyne.

‡ Formerly with the Department of Electrical Engineering, The University of Newcastle upon Tyne; now with the Department of Engineering Science, University of Durham.

used for shifting a given zero. With this number of arbitrary choices it is obvious that the number of different realizations is indeed large.

The question now arising is: which of the possible realizations should be selected? Various factors might influence a decision, e.g. the number of components, the component spread, the circuit configuration and the number of reactive and resistive components could be of interest. But also the constant multiplier or gain constant H associated with every transfer function could be of importance. The gain constant H has a maximum value for R-C realization procedures and this value is considered here. It is important to note that since one is interested in the problem of what can be obtained from passive networks alone, the gain constant H is as essential as the numerator and denominator polynomials of a transfer function. Furthermore, if it is possible to obtain networks giving maximum gain, the noise level compared with the output signal level will be lower. If, apart from maximum gain, a reduction in the number of network elements is obtained with a certain synthesis procedure, then this is clearly an extremely useful approach which should be duly considered.

The maximum realizable gain of an R-C two-port has been studied extensively by Failkow and Gerst.[5-7] Cederbaum[8] has derived results for special cases where transmission zeros are restricted to the origin and infinity. There the upper limit of the realizable gain is derived in terms of the given poles and zeros. A paper relating to this specific problem was presented by Paige and Kuh.[9] The maximum value which H can attain is determined by the nature of the numerator and denominator polynomial of a transfer function in conjunction with the R-C realizability conditions. The maximum value for H can be obtained graphically by using the Bode plot where $\log_e T_{21}(j\omega)$, the transfer function, is plotted against $\log_e \omega$. Given a transfer function of the type

$$T_{21}(p) = \frac{E_2}{E_1} = H\frac{\prod_{j=1}^{n}(p+z_j)}{\prod_{i=1}^{m}(p+p_i)} \qquad \ldots\ldots(1)$$

the maximum gain for a low-pass type transfer function is given by

$$H_{\max} = \frac{\prod_{1}^{m} p_i}{\prod_{1}^{n} z_j} \qquad \ldots\ldots(2)$$

The use of this equation ensures that for a low-pass transfer function the frequency response will start from the 0 dB line. For a high-pass transfer function the highest voltage gain is taken to be at infinite frequency and therefore

$$H_{\max} = 1 \qquad \ldots\ldots(3)$$

For the band-pass case $H_{\max}$ can be approached as closely as desired. It is obvious that neither at d.c. nor at $\omega = \infty$ does maximum gain occur. Therefore, the synthesis procedure depends on the splitting of the transfer function or at least the band-bass network into a cascade of low-pass and high-pass ladders. It is in this splitting process that one discovers that $H_{\max}$ cannot be achieved.

3. The Basic Requirements for the Synthesis

The starting point for the synthesis is prescribed by the voltage transfer function $T_{21}(p)$ expressed by

$$T_{21}(p) = \frac{-Y_{21}}{Y_{22}} = H\frac{N(p)}{D(p)} \qquad \ldots\ldots(4)$$

Now if the transmission zeros of $T_{21}(p)$ do not all lie on the negative real axis, it is known that a single ladder realization is not possible. However, by splitting up the Y_{21} parameters in such a way that the transmission zeros of each all lie on the negative real axis, the realization of individual ladders is made possible. It is required that the numerator coefficients of $-Y_{21}$ be positive. The complete voltage transfer function for a parallel ladder network may then be given as

$$T_{21}(p) = \frac{-(Y_{21a}+Y_{21b}+Y_{21c}+\ldots+Y_{21n})}{Y_{22a}+Y_{22b}+Y_{22c}+\ldots+Y_{22n}} \qquad \ldots\ldots(5)$$

where the suffixes $a, b, c\ldots$, etc., refer to the parameters of particular ladder networks. In some cases it may be necessary to apply zero shifting techniques in order to realize the required transmission zero.

Now pursuing the idea of the parallel ladder realization one would expect the given transfer function to be a ratio of two polynomials in the following way

$$T_{21}(p) = H\frac{p^n+a_1p^{(n-1)}+a_2p^{(n-2)}+\ldots+a_n}{p^m+b_1p^{(m-1)}+b_2p^{(m-2)}+\ldots+b_m} \qquad \ldots\ldots(6)$$

where all the coefficients are positive real. For the numerator polynomial of eqn. (6) two cases can arise; either the zeros can be negative real or they may be complex with the special case when they lie on the $j\omega$-axis. The poles of eqn. (6) for passive realizability must be negative real. If $N(p)$, the numerator polynomial has right half-plane zeros, its coefficients may still be positive. Therefore so long as $N(p)$ does not have zeros on the positive real axis, one can obtain a resultant polynomial $N(p)$ with positive coefficients by augmenting $T_{21}(p)$ with factors of the type $(p+a)$.

The number of parallel ladders required for a given transfer function depends on the nature of the numerator polynomial, thus, if

$$N(p) = \prod_{i=1}^{n} K_i(p+z_i); \quad z_i \text{ real} \quad \ldots\ldots(7)$$

the realization will yield a simple ladder network using zero shifting techniques. For

$$N(p) = \prod_{i=1}^{n} K_i(p^2+z_i^2); \quad z_i \text{ real} \quad \ldots\ldots(8)$$

the zeros are all on the $j\omega$-axis, i.e. only even-order terms exist. For this, however, the same number of parallel ladders is realized as for the numerator polynomial of eqn. (6) where the zeros are all complex.

Thus the number of parallel ladders required for a general numerator polynomial as in eqn. (6) is given by

$$q = \frac{2n+3+(-1)^n}{4} \quad \ldots\ldots(9)$$

where n is the degree of the numerator polynomial. Furthermore, for the realization of eqn. (6) a divisor polynomial must be introduced. It is the order of this divisor polynomial that eventually limits the component economy of the realized network. Because $m \geqslant n$ in eqn. (6), the choice of the divisor polynomial is made with reference to the denominator polynomial. To test the driving point admittances for the positive real character of rational polynomial quotients requires that the following necessary conditions be fulfilled:

(a) all polynomial coefficients be real and positive;

(b) degrees of numerator and denominator polynomials differ at most by one;

(c) numerator and denominator terms of lowest degree differ at most by one;

(d) imaginary axis poles and zeros be simple;

(e) there be no terms missing in numerator and denominator polynomials unless all even or all odd terms are missing.

Let the divisor polynomial $F(p)$ be chosen one degree lower than the denominator polynomial eqn. (6). It must be remembered that for R-C impedance or admittance functions the poles and zeros interlace and for an admittance function the lowest critical frequency is a zero which may be at $p = 0$. As the zeros of the denominator polynomial in eqn. (6) are simple and located on the negative real axis of the p-plane it seems relatively easy to arrange for the zeros of $F(p)$ to interlace with those of $D(p)$, the denominator polynomial. This is true in general, but in the synthesis proposed here two aims are to be fulfilled:

(i) the realized network should be a maximum gain structure;

(ii) a minimum number of elements are to be realized.

This now requires that the divisor polynomial be selected after the zeros of the transfer admittances Y_{21i}, and the driving point admittances Y_{22i} are known. For example, in the Guillemin synthesis procedure the numerator polynomials for all Y_{22} would be the same and then impedance scaling would be applied after the individual ladders were derived and before they are assembled to form the complete network. Here, however, it is proposed to split the denominator polynomial of eqn. (6) into sub-polynomials and it is then required that the sub-polynomials have zeros interlacing with those of the divisor polynomial. Thus the driving point function would be

$$Y(p) = \frac{\prod_{i=1}^{l} K_i(p+p_i)}{\prod_{j=1}^{0} (p+d_j)} \quad \ldots\ldots(10)$$

with

$$\frac{p_i}{K_i} = d_j \text{ for } (i,j) = 1,2,3\ldots.$$

The problem of synthesizing a transfer function having only negative real zeros is relatively simple. A ladder network evolves which will have minimum number of components if the degree of the divisor polynomial is one order lower than the denominator polynomial $D(p)$. Maximum gain is ensured if H is chosen correctly after identifying the function. The number of components which will be realized from a certain given transfer function depends primarily on the number of terms, k, in the driving point admittances Y_{22i} but also to a certain extent on the transfer admittances Y_{21i}, because they dictate the number of zero shifts z_s which must be performed. Thus the number of components, N, realized from a driving point admittance Y_{22i} and its transfer admittance Y_{21i} is given by:

$$N = \left(k-1+\sum_{s=1}^{l} z_s\right)+\tau \quad \ldots\ldots(11)$$

$$\tau = 0 \text{ when } a_n < b_m$$

$$\tau = 1 \text{ when } a_n > b_m, a_{n-1} \neq 0$$

It is clear that when the number of ladders of a transfer function has been established, the number of components can also be determined. It will now be shown how the final network is derived.

For eqn. (6) to represent a low-pass function $a_n > b_m$. Assuming all the zeros to be complex, the function may be split in the following way:

$$\left.\begin{aligned}
&Y_{21a} = \frac{p^n + a_1 p^{n-1}}{F(p)} && Y_{22a} = \frac{p^m + c_a p^{m-1} + d_a p^{m-2} + \ldots + y_a p}{F(p)} \\
&Y_{21b} = \frac{a_2 p^{n-2} + a_3 p^{n-3}}{F(p)} && Y_{22b} = \frac{c_b p^{m-1} + d_b p^{m-2} + \ldots + y_b p}{F(p)} \\
&\quad\vdots && \quad\vdots \\
&Y_{21i} = \frac{a_{n-1} p + a_n}{F(p)} && Y_{22i} = \frac{\left(\frac{1}{H} - 1\right)p^m + \left(b_1 - \sum_a^{i-1} c_r\right)p^{m-1} + \left(b_2 - \sum_a^{i-1} d_r\right)p^{m-2} + \ldots + b_m}{F(p)}
\end{aligned}\right\} \quad \ldots\ldots(12)$$

If $m = n$, then the first and the last ladder in eqn. (12) will be a high-pass and a low-pass ladder respectively. The other ladders exhibit band-pass or band-stop characteristics depending on the nature of Y_{21i}. $F(p)$, the divisor polynomial is of order $(m-1)$, its zeros are required to interlace with those of all the Y_{22}. This requires that all the zeros of the Y_{22} are known in factorized form. Alternatively, one may choose the coefficients c_r, d_r, etc., to be so small that the locations of the zeros of $D(p)$ are not changed by much.

$$T_{21}(p) = H \frac{(p^2+9)(p^2+4)}{(p^2+4p+0\cdot9)(p^2+3p+0\cdot4)} \quad \ldots\ldots(13)$$

This is a low-pass function, thus $H = 0\cdot01$ with the zeros of the denominator polynomial situated at $p_1 = -0\cdot14$, $p_2 = -0\cdot254$, $p_3 = -2\cdot86$ and $p_4 = -3\cdot76$. These pole-locations, however, cannot be used to determine the choice of divisor polynomial. Therefore it is necessary to first establish the admittance parameters of the individual ladders. According to eqn. (9), $q = 3$ and therefore three parallel ladders are required:

$$\left.\begin{aligned}
&Y_{21a} = \frac{p^4}{F(p)} && Y_{22a} = \frac{p(p+0\cdot5)(p+2\cdot2)(p+3\cdot5)}{F(p)} \\
&Y_{21b} = \frac{13p^2}{F(p)} && Y_{22b} = \frac{4\cdot8148p(p+0\cdot6)(p+2\cdot1)}{F(p)} \\
&Y_{21c} = \frac{36}{F(p)} && Y_{22c} = \frac{99p^4 + 688\cdot98p^3 + 1306\cdot45p^2 + 428\cdot08p + 36}{F(p)}
\end{aligned}\right\} \quad \ldots\ldots(14)$$

Attention must be paid to the ladders b to $(i-1)$; that is, all those except the first and the last, as they may be required to be impedance scaled in order to obtain the proper transfer admittance Y_{21}. Once the individual ladders have been derived, its Y_{21} should agree with that stipulated by the transfer function; if this is not so, scaling of Y_{21} will be necessary.

From eqn. (12) it is evident that H has been made operative on $D(p)$, the denominator polynomial. Now a number of different polynomial configurations are possible with given transfer functions. Thus if $m \geq n$, all odd-order terms $(2r-1)$, $r = 1, 2, 3$, etc., of the numerator polynomial may be missing. Any such reduction in polynomial terms automatically reduces the number of elements realized.

4. Example of a Fourth-order Rational Transfer Function

The transfer function to be synthesized is

The zeros of Y_{22a} and Y_{22b} are chosen to be close to those of $D(p)$ with the aim of keeping the coefficients in the numerator terms of the driving point admittances as small as possible. If this can be achieved then the zeros of Y_{22c} will not differ appreciably from those of $D(p)$. The factor 4·8148 in Y_{22b} was introduced so as to make the second-order coefficient equal to 13, i.e. the same as in Y_{21b}. This is done in order to escape scaling of the impedances after realization. However, it will be seen later that this technique is superfluous.

The divisor polynomial must now be chosen. All the zeros of the Y_{22} in eqn. (14) are known and those of $F(p)$ are selected so as to interlace with them. An appropriate choice would be

$$F(p) = (p+0\cdot2)(p+2)(p+3\cdot3) \quad \ldots\ldots(15)$$

The number of components which are expected to be realized are $N_a = 7$ for Y_{22a}, $N_b = 6$ for Y_{22b} and $N_c = 8$ for Y_{22c}.

Applying the continued fraction method described by Cauer,[10] Y_{22a}, Y_{22b} and Y_{22c} are realized as depicted in Fig. 1(a), (b) and (c) respectively.

An analysis of Fig. 1(b) shows that

$$Y_{21b} = \frac{0{\cdot}1400499p^2}{F(p)};$$

however, according to eqn. (14) it is required to be

$$Y_{21b} = \frac{13p^2}{F(p)}$$

This, therefore, is a case where it is necessary for the admittance to be scaled. The multiplying factor between the realized and the required Y_{21b} is found to be 92·8538.

As the numerator of Y_{21b} consists of the product of four admittances, the scaling factor for the individual admittances will be given by

$$\sqrt[4]{92{\cdot}8538} = 3{\cdot}1042$$

Thus the ladder of Fig. 1(b) is now replaced by Fig. 1(d), and the complete circuit then is the parallel connection of the ladders (a), (c) and (d) of Fig. 1. All components are normalized and the values given are in ohms and farads.

If the Guillemin parallel ladder synthesis were applied to the transfer function eqn. (13), then a circuit containing 27 or perhaps 24 elements would result instead of the 21 which are realized here.

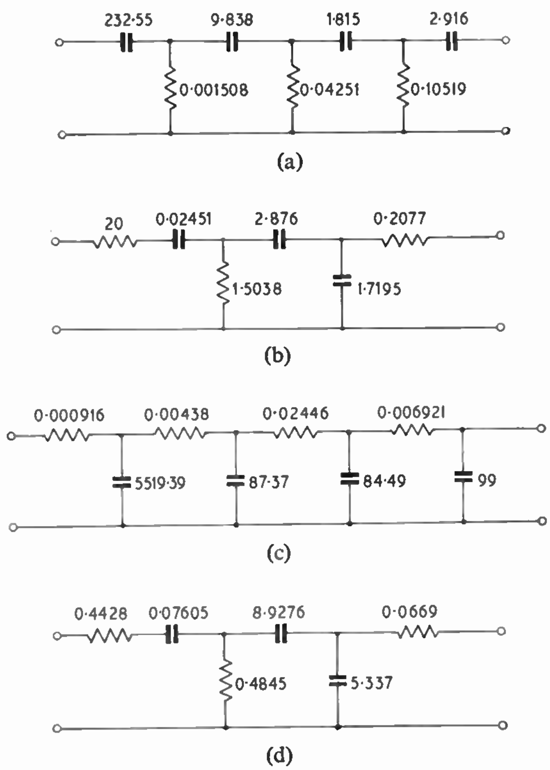


Fig. 1. Parallel ladders of a fourth-order network.

5. Augmented Function to Realize Right-half Plane Zeros

Another example in the synthesis of a passive R-C network is given. The transfer function here used is taken from a practical application in active R-C synthesis of a second-order all-pass function which was realized using an operational amplifier as the active element.

The original function is

$$T_{21}'(p) = H\frac{p^2-2p+10}{p^2+11p+5} \qquad \text{......(16)}$$

It is, of course, possible to use an active synthesis procedure for the realization of this function. However, a passive network should be preferred both from the point of stability and sensitivity to component changes.

Now using a common factor $(p+4)$ for both numerator and denominator polynomial of eqn. (16) results in a new function

$$G_{21}(p) = H\frac{p^3+2p^2+2p+40}{(p+0{\cdot}4706)(p+4)(p+10{\cdot}52494)} \qquad \text{......(17)}$$

The two transfer functions are of the low-pass kind so that $H = 0{\cdot}5$ and with this the admittance parameters are given by:

$$\left.\begin{aligned} Y_{21a} &= \frac{p^3+2p^2}{F(p)} \qquad Y_{22a} = \frac{p(p+4)(p+8)}{F(p)} \\ Y_{21b} &= \frac{2p+10}{F(p)} \qquad Y_{22b} = \frac{(p+0{\cdot}820916)(p+4{\cdot}002256)(p+12{\cdot}194433)}{F(p)} \end{aligned}\right\} \qquad \text{......(18)}$$

The divisor polynomial $F(p)$ can now be chosen to interlace its zero with those of Y_{22a} and Y_{22b}; a convenient choice being

$$F(p) = (p+3)(p+8) \qquad \text{......(19)}$$

In both cases for ladder (a) and (b) the zero-producing step is left till the end of the ladder development. The complete network is depicted in Fig. 2. Compared with the Guillemin method used by Sewell,[11] a saving of three components has been achieved in addition to

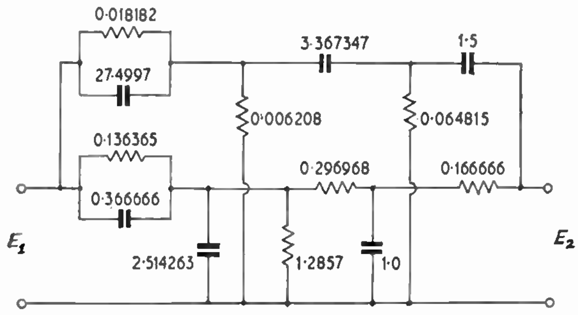


Fig. 2. Third-order rational transfer function network.

a lower component value spread and maximum gain. It is thought that the method presented here may be used to great advantage in numerous applications.

6. Conclusions

The aim of this paper has been to show how it is possible to reduce the number of circuit elements and yet maintain an overall maximum gain at all frequencies. There are undoubtedly more economical circuits but these do not produce maximum gain. It may therefore be said that for the parallel ladder arrangement as it is proposed here, a minimum number of components is obtained.

The advantage of this method is due to its relative simplicity and the ease with which the final number of parallel ladders and components can be calculated.

It is hoped that this contribution may provide for some useful practical applications and stimulate new thought and ideas for other researchers in this field.

7. References

1. Guillemin, E. A., 'Synthesis of Passive Networks' (Wiley, New York, 1957; Chapman & Hall, London, 1959).
2. Guillemin, E. A., 'Synthesis of RC networks', *J. Math. Phys.*, **28**, pp. 22–42, 1949.
3. Dasher, B. J., 'Synthesis of RC transfer functions as unbalanced two terminal-pair networks', *Trans. Inst. Radio Engrs on Circuit Theory*, **CT-1**, pp. 20–34, December 1952.
4. Holt, A. G. J. and Reineck, K. M., 'The transfer function synthesis for a cascade connection network', *Trans. Inst. Elect. Electronics Engrs on Circuit Theory*, **CT-15**, pp. 162–5, June 1968.
5. Failkow, A. D., 'Two terminal-pair networks containing two kinds of elements only', Proc. Symposium Modern Network Synthesis, pp. 50–65, 1952.
6. Failkow, A. D. and Gerst, J., 'The transfer function of an RC ladder network', *J. Math. Phys.*, **30**, pp. 49–71, 1952.
7. Failkow, A. D. and Gerst, J., 'The transfer function of general two terminal-pair RC networks', *Quart. Appl. Math.*, **10**, No. 2, pp. 113–27, 1952.
8. Cederbaum, I., 'Some properties of transfer functions of unbalanced RC networks', *Proc. Instn Elect. Engrs*, **103-C**, pp. 400–406, 1956. (I.E.E. Monograph No. 180R, June 1956.)
9. Paige, A. and Kuh, K. S., 'Maximum gain realization of an RC ladder network', *Trans. Inst. Radio Engrs on Circuit Theory*, **CT-9**, No. 1, pp. 32–40, 1960.
10. Cauer, W., 'Synthesis of Linear Communications Networks', Vols. 1 and 2 (McGraw-Hill, New York and London, 1958.)
11. Sewell, J. I., 'The Application of Analogue Techniques to the Design of Active Filters', Ph.D. Thesis, University of Newcastle upon Tyne, October 1966.

Manuscript first received by the Institution on 5th April 1968 and in final form on 24th July 1968. (Paper No. 1220/CC26.)

U.D.C. 621.372.622:621.382.2

High-frequency Mixers using Square-law Diodes

By
Professor
D. P. HOWSON, D.Sc., C.Eng., F.I.E.E., M.I.E.R.E.†
AND
J. G. GARDINER, B.Sc., Ph.D., (*Graduate*)‡

Summary: The use of symmetrical and asymmetrical space charge limited (s.c.l.) square-law diodes as mixer elements is considered, and it is shown that low conversion loss is possible with the latter type. In order to achieve this a d.c. bias is necessary on the diode. Matching source and load resistances are higher than for the unbiased diode. The use of an idler termination at the sum frequency ($p+q$) is shown to have some advantages. The symmetrical diode is shown to be unpromising as a mixer element.

List of Symbols

A_1	conversion power loss ratio, narrow-band mixer, image short-circuit
A_2	conversion power loss ratio, narrow-band mixer, image open-circuit
A_3	conversion power loss ratio, broad-band mixer, termination identical for signal and image
g_b	conductance of reverse-biased asymmetrical diode
$g(t)$	time-variation of diode conductance
g_0	d.c. component of $g(t)$
g_1	component of $g(t)$ at local-oscillator frequency
g_n	component of $g(t)$ at nth harmonic of local-oscillator frequency
i	diode current
k	coefficient of v^2 term in diode d.c. characteristic
l	coefficient of linear term in diode d.c. characteristic
n	order of local-oscillator harmonic
p	angular frequency of local oscillator
q	angular frequency of received signal
$p-q$	angular frequency of wanted modulation product
$q-2p$	angular frequency of image
Q_0	diode quality factor referred to d.c. characteristic
v	voltage appearing across diode
v_0	diode d.c. bias voltage
$2V$	amplitude of local-oscillator component of diode voltage
α/π	fraction of local-oscillator period during which diode is forward biased
$\theta = pt$	

† Formerly in the Department of Electronic and Electrical Engineering, University of Birmingham; now in the Schools of Electrical and Electronic Engineering, University of Bradford.

‡ Formerly in the Department of Electronic and Electrical Engineering, University of Birmingham; now with Racal Communications Ltd., Tewkesbury, Gloucestershire.

1. Introduction

The recent introduction of hot-carrier diodes as mixer elements has resulted in a significant reduction in the noise figure of systems using a mixer as the first stage.[1] Although at present conversion losses of 5 dB and system noise figures of 6·5 dB are being quoted for commercial hot-carrier mixers using a following 1·3 dB i.f. amplifier, laboratory mixers have produced results rather better than this. Indeed losses as low as 1·3 dB have been quoted in some cases, with corresponding system noise figures of 2·7 dB.[1]

Good as the hot-carrier diode mixer undoubtedly is, however, it does have some disadvantages. Improved intermodulation performance in mixers has been suggested, when square-law diodes are substituted for the hot-carrier diodes.[2, 3]

Recent papers[4, 5] have discussed the noise in space-charge-limited square-law diodes, and there has been speculation about the possibility of very low noise devices, due to the shot noise being space-charge smoothed. Measurements have so far failed to confirm the low noise properties of the diodes, the limit appearing to be the thermal noise of the incremental conductance.[5] However, the matter does not appear to be finally resolved, and there remains the possibility of using such diodes in a mixer to obtain very low values of excess noise and, therefore, if the mixer conversion loss can be sufficiently low, a correspondingly low system noise figure.

For these reasons it was decided to investigate whether low-loss mixers using square-law diodes were practicable, for both forms of diode under construction at present.

2. The Space-charge Limited Diode

The structure of electron-injecting diodes of this type is shown in Fig. 1. The devices are essentially layers of high resistivity semiconductor, usually silicon,[6] containing very few thermally-generated carriers at room temperature, connected to the external circuit by source and drain contacts.

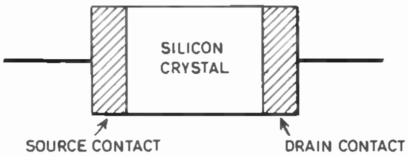


Fig. 1. Structure of s.c.l. diodes.

In practice the high-resistivity layer may for convenience be epitaxially grown on a low-resistivity substrate.

From the source contact, which may be a gold/antimony alloy, mobile carriers may be drawn into the crystal under conditions of forward voltage bias. A space-charge cloud of mobile carriers forms in the silicon close to the source contact, and current under this condition consists of carriers drawn from this cloud through the silicon to the drain contact. This latter contact may either be identical to the source contact, in which case the diode is said to be symmetrical, or be a blocking contact, usually gold, in which case the diode is asymmetrical.

The current-voltage relationship for forward bias is governed by the fact that the amount of mobile charge between the contacts, and also the charge velocity, are proportional to the applied voltage. The diode current i may be shown theoretically to be proportional to the square of the applied voltage v, i.e.

$$i = kv^2 \qquad \ldots\ldots(1)$$

where k is a constant determined by the device geometry, the dielectric constant of the silicon and the carrier mobility.

For small signals it is found that deviations from this law in practical diodes can be accounted for by the addition of linear shunt conductance l, so that

$$i = kv^2 + lv \qquad \ldots\ldots(2)$$

Some recent calculations by Wright[6] have indicated that a diode with a source-to-drain spacing of 1 μm will make an efficient microwave detector over the centimetre and millimetre bands. Analysis of operation as a mixer diode is more complex but significant deviations from eqn. (2) would not be expected below microwave frequencies.

Characteristics of the two forms of s.c.l. diode are shown in Fig. 2, from which it can be seen that the asymmetric diode becomes a high impedance when reverse biased, since the blocking drain contact cannot inject mobile charge carriers into the silicon. Experimental work by Wright and his associates has suggested that this form of diode is a little more difficult to fabricate successfully than the symmetrical type, which will accordingly be considered first.

3. Mixer Analysis using Symmetrical Diodes

From the form of the diode characteristic given in Fig. 2, it can be seen that the current is

$$i = kv^2 \operatorname{sign}(v) \qquad \ldots\ldots(3)$$

ignoring, for the moment, any linear term.

The incremental conductance is

$$\frac{\mathrm{d}i}{\mathrm{d}v} = 2kv \operatorname{sign}(v) = 2k|v| \qquad \ldots\ldots(4)$$

The mixer can be analysed as a linear, time-varying circuit for the signal and the modulation products as long as it can be assumed that the pump voltage alone controls the incremental conductance of the diode. This is equivalent to the assumption that the pump

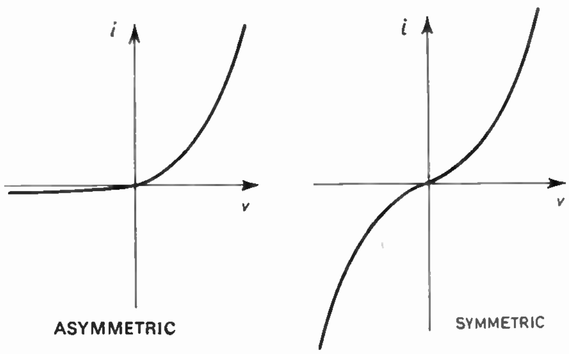


Fig. 2. Diode characteristics.

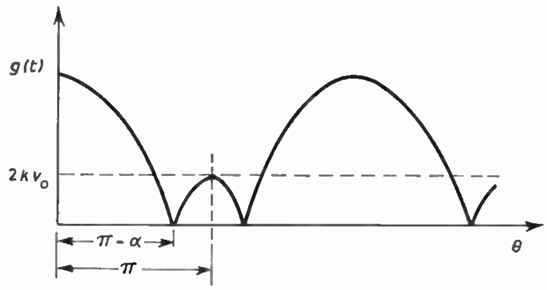
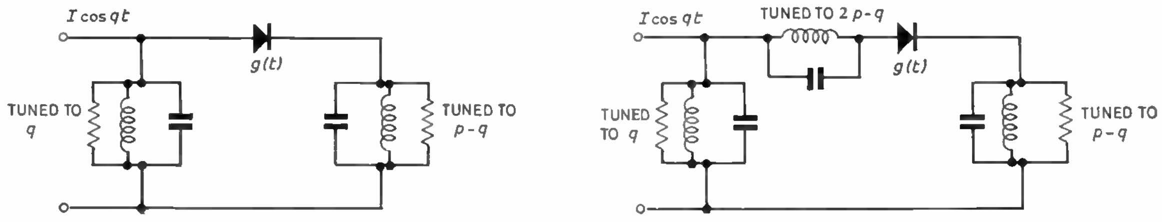


Fig. 3. Symmetrical diode conductance. ($v_0 = V$).



(a) Image short-circuited. (b) Image open-circuited.

Fig. 4. Mixer circuits.

voltage across the diode is very much larger than the corresponding signal voltage. If we consider the diode to be pumped with a sinusoidal voltage of amplitude $2V$ and allow in addition a d.c. bias, v_0, then the voltage across the diode is

$$v = v_0 + 2V \cos pt \qquad \text{......(5)}$$

and

$$g(t) = \frac{di}{dv} = 2k|v_0 + 2V \cos pt| \qquad \text{......(6)}$$

In general $g(t)$ is of the form shown in Fig. 3, where $\theta = pt$ appearing in the form of a full-wave rectified sinusoid for $v_0 = 0$, and also for $v_0 = 2V$. The range of values of biasing voltage of primary interest lies between these two points, since if $v_0 \geqslant 2V$ reverse bias is never established across the diode so that the symmetry of the diode i/v characteristic about the origin is not exploited.

Since $g(t)$ is a periodic function of time with fundamental frequency p, it may be written

$$g(t) = g_0 + 2\sum_{n=1}^{\infty} g_n \cos npt \qquad \text{......(7)}$$

Analysis of the waveform of Fig. 3 gives

$$g_0 = \frac{4kV}{\pi}\{2 \sin \alpha - 2\alpha \cos \alpha + \pi \cos \alpha\} \qquad \text{......(8)}$$

$$g_1 = \frac{2kV}{\pi}\{2\alpha - \sin 2\alpha - \pi\} \qquad \text{......(9)}$$

$$g_2 = \frac{4kV}{3\pi}\{\sin 2\alpha \cos \alpha - 2 \cos 2\alpha \sin \alpha\} \qquad \text{......(10)}$$

where α is defined as $\cos^{-1}(v_0/2V)$.

Using Strum's analysis,[7] the minimum conversion loss of a mixer with this diode may now be calculated, for the image frequency, $q-2p$, either open or short-circuited. Figure 4 gives circuits for the two types of mixer.

For the image frequency short-circuited, the minimum conversion loss is given by

$$A_1 = \left\{\frac{1+\sqrt{1-g_1^2}}{g_1}\right\}^2 \qquad \text{......(11)}$$

and for the image frequency open-circuited by

$$A_2 = \left\{1 + \sqrt{\frac{1+g_2-2g_1^2}{(1-g_1^2)(1+g_2)}}\right\}^2 \left\{\frac{(1-g_1^2)(1+g_2)}{g_1^2(1-g_2)}\right\} \qquad \text{......(12)}$$

In eqns. (11) and (12) g_1 and g_2 are normalized by division with g_0. The resultant conversion loss in each case was calculated, and is plotted in Fig. 5, as a function of α. Figure 6 gives the optimum conversion loss for practical diodes which have, in addition to the square-law term, a shunt conductance. It can be seen that the loss cannot be reduced below 10·4 dB, and that the addition of shunt conductance worsens this result.

Considering now the case when $v_0 \geqslant 2V$, so that the applied voltage is always positive, we have

$$g(t) = 2k(v_0 + 2V \cos pt) \qquad \text{......(13)}$$

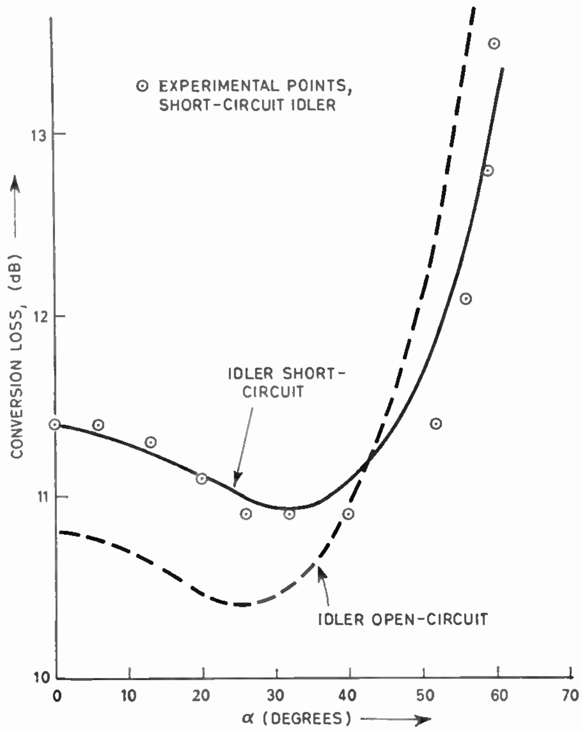


Fig. 5. Symmetrical diode—zero linear conductance.

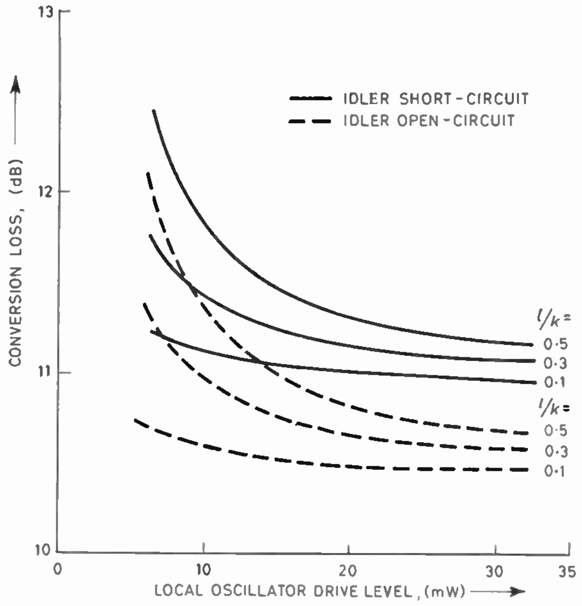


Fig. 6. Influence of shunt linear conductance on loss.

and, therefore, using Fourier analysis,

$$g_0 = 2kv_0 \qquad \text{......(14)}$$

$$g_1 = 2kV \qquad \text{......(15)}$$

$$g_2 = 0 \qquad \text{......(16)}$$

Substitution of these values into the conversion loss equations shows that the minimum loss for the image frequency short-circuited is 11·4 dB, and for the image frequency open-circuited is 10·8 dB. None of the ways of using the symmetrical diode that have been analysed give a mixer conversion loss low enough to be attractive, certainly below 30 GHz.

4. Mixer Analysis using Asymmetrical Diodes

For this type of diode the conductance for negative bias was considered to be constant. The i–v characteristic of the diode was taken as

$$i = kv^2 + g_b v \qquad v > 0 \qquad \text{......(17)}$$

$$i = g_b v \qquad v < 0 \qquad \text{......(18)}$$

The coefficient of the linear term for positive v was chosen to avoid a discontinuity in $g(t)$ at the origin (see Fig. 7). Taking the same form of expression for the applied voltage as before with a changed sign for the bias voltage, v_0,

$$g(t) = \frac{di}{dv} = 2k(2V \cos pt - v_0) + g_b \qquad v > 0 \qquad \text{......(19)}$$

$$= g_b \qquad v < 0 \qquad \text{......(20)}$$

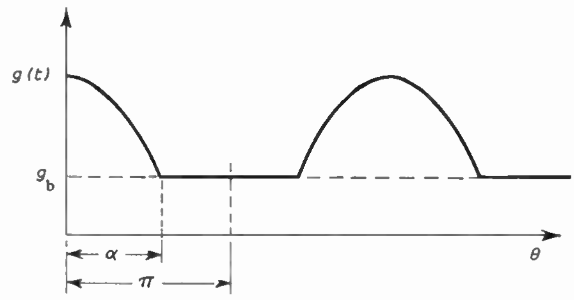


Fig. 7. Asymmetrical diode conductance.

By Fourier analysis, it is found that

$$g_0 = \frac{4Vk}{\pi}(\sin\alpha - \alpha\cos\alpha) + g_b \qquad \text{......(21)}$$

$$g_1 = \frac{2Vk}{\pi}\left(\alpha - \frac{\sin 2\alpha}{2}\right) \qquad \text{......(22)}$$

$$g_n = \frac{4Vk}{\pi}\left[\frac{\sin n\alpha\cos\alpha - n\cos n\alpha\sin\alpha}{n(n^2-1)}\right] \quad \text{for} \quad n > 1 \qquad \text{......(23)}$$

Define a diode quality factor of

$$Q_0 = \frac{kV}{g_b} \qquad \text{......(24)}$$

Figure 8 gives the minimum conversion loss of a mixer using a biased asymmetrical diode, as a function of α for two values of Q_0.

It is estimated that practical diodes should lie within this range of Q_0. The reverse conductance is small; the device behaves in reverse bias in exactly the same way as a conventional Schottky barrier diode, whilst it is considered that suitable choice of device geometry can

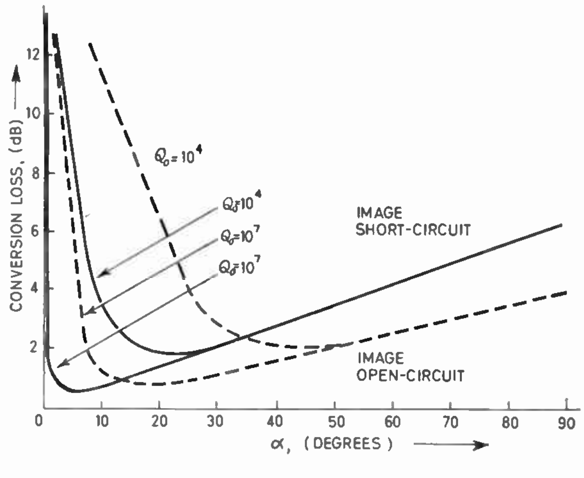


Fig. 8. onversion loss for narrow-band mixer using asymmetric diode.

result in a high forward conductance without necessarily introducing a large shunt capacitance.

It is assumed here that the image frequency is open- or short-circuited, i.e. that it is a narrow-band mixer. It can be seen that for diodes with a high quality factor, losses less than 1 dB are obtainable.

It should be noted, however, that the matching resistances for source and load increase in value as α decreases, as shown in Fig. 9. It may be impracticable, therefore, with a particular diode, to decrease α sufficiently to obtain minimum loss because of the high value of matching resistances required.

In many applications a broad-band mixer is required, in which the termination at the image frequency is the same as that at the signal frequency. In this case the minimum conversion loss is given by

$$A_3 = \left\{1 + \sqrt{\frac{1+g_2-2g_1^2}{1+g_2}}\right\}^2 \left\{\frac{1+g_2}{g_1^2}\right\} \quad \ldots\ldots(25)$$

where g_1 and g_2 are normalized by division with g_0, according to Strum.[7] Substitution of the coefficients for the biased square-law diode results in the curve of Fig. 10. From this it can be seen that for diodes with a high quality factor, losses less than 3·5 dB are possible, i.e. losses only 0·5 dB above the absolute minimum for broad-band operation.

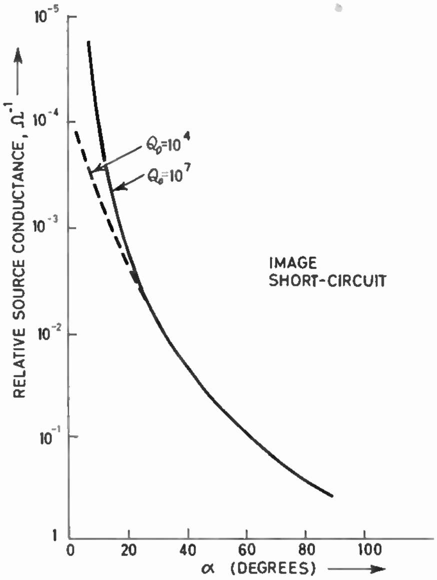


Fig. 9. Source conductance for narrow-band mixer using asymmetric diode.

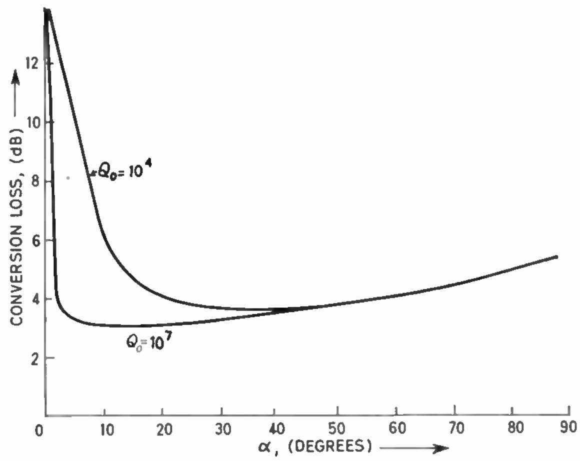


Fig. 10. Conversion loss of broad-band mixer using asymmetrical diode.

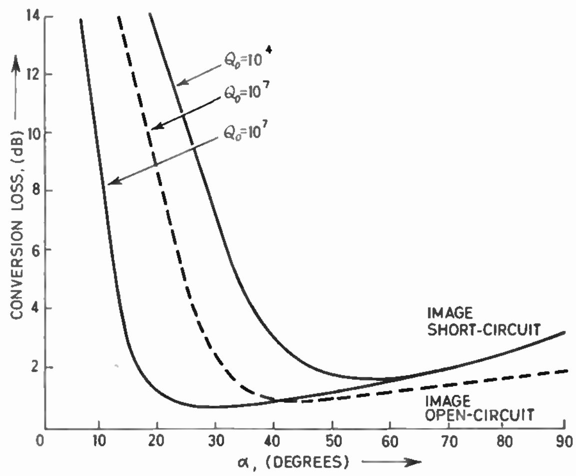


Fig. 11. Conversion loss for narrow-band mixer, using asymmetric diode and additional idler circuit at angular frequencies of $p+q$, $3p-q$ rad/s.

4.1. *Use of an Additional Idler Circuit*

It has been shown previously[7] that it is possible to decrease the loss in a mixer by the correct termination of the first upper sideband, $p+q$. The present authors consider that an analysis based on this fact may well be misleading, since another sideband, $3p-q$, is very close in frequency to $p+q$ so that any termination of one may well affect the other. Preliminary calculations have shown that ignoring this latter modulation product can lead to errors of up to 0·5 dB. Accordingly, an analysis was made using a method developed earlier[8] for work on rectifier modulators. Using the conductance coefficients calculated in eqns. (21)–(23), the minimum conversion loss of a mixer for which both $p+q$ and $3p-q$ were open-circuited was derived. The results are shown in Fig. 11. It can be seen that this modified mixer has considerably lower loss for large values of α, particularly for diodes with a high quality factor, but that the minimum obtainable loss for such a diode was not reduced. However, a useful reduction in minimum loss can be seen for relatively poor diodes. In some cases the matching resistances for a particular loss are lower than in the simpler circuit (see Fig. 9).

5. Conclusions

It has been shown that a mixer using a symmetrical square-law diode cannot be designed for an acceptable low loss using conventional techniques. In contrast, the asymmetrical square-law diode, if d.c. bias is provided, can be used in a mixer to give a conversion loss of less than 1 dB. In order to obtain this result the mixer must present an open-circuit to the image frequency, and also the square-law diode must have a high quality factor, where this is defined as the ratio of the diode constant k multiplied by the peak applied alternating voltage, all divided by the diode reverse conductance. If such a diode is used in a broadband mixer, where both signal and image frequencies share the same termination, a minimum loss approaching 3·5 dB appears attainable.

In conclusion, the effect on the conversion loss of the narrow-band mixer using an asymmetric square-law diode of the termination of the upper sideband, $p+q$, is considered. It is shown that if this frequency, together with its near neighbour $3p-q$, is open-circuited, lower loss may be obtained under certain conditions.

6. References

1. 'Hot carrier diodes—their design and applications', *Electronic Components*, p. 527, June 1965.
2. Gardiner, J. G., 'Cross modulation and intermodulation distortions in the tuned square-law diode modulator', Memorandum No. 285. Electronic and Electrical Engineering Department, University of Birmingham.
3. Lepoff, J. H. and Cowley, A. M., 'Improved intermodulation rejection in mixers', *Trans. Inst. Elect. Electronics Engrs on Microwave Theory and Techniques*, **MTT-14**, pp. 618–23, December 1966.
4. Wright, G. T., 'Space-charge-limited solid-state devices', *Proc. I.E.E.E.*, **51**, p. 1642, November 1963.
5. Hsu, S. T., Van der Ziel, A. and Chennette, E. R., 'Noise in s.c.l. solid state devices', *Solid State Electronics*, **10**, p. 129, February 1967.
6. Wright, G. T., 'Transit time effects in the s.c.l. silicon microwave diode', *Solid-State Electronics*, **9**, p. 1, January 1966.
7. Strum, P. D., 'Some aspects of crystal mixer performance', *Proc. I.R.E.*, **41**, p. 875, July 1953.
8. Howson, D. P., 'Low-loss shunt or series rectifier modulators', *Electronics Letters*, **2**, p. 225, June 1966.

Manuscript first received by the Institution on 19th March 1968 and in final form on 19th July 1968. (Paper No. 1221/CC27.)

U.D.C. 629.19:621.396.61

A Satellite-borne Receiver for Low-frequency Radio Astronomy

By

D. WEIGHTON,
M.A., C.Eng., M.I.E.E.†

Presented at a Symposium on 'The Ariel III Satellite' held in London on 13th October 1967.

Summary: A receiver designed to measure the galactic noise spectrum in the range from 2 to 4·5 MHz and flown in the *Ariel III* satellite is described. Problems arising from the use of a loop antenna with a swept receiver and the exclusion of spurious signals from terrestrial sources and other experiments in the payload are discussed. Preliminary results indicate a marked coupling between the receiver and another experiment in the satellite by modulation of the local plasma. In certain conditions the coupling is small enough to be neglected and the galactic noise level can be measured with fair accuracy.

1. Introduction

This paper describes the design of the equipment used in an experiment flown by the University of Manchester (Jodrell Bank) in the *Ariel III* satellite. The purpose of the experiment is to explore the spectrum of galactic noise in the band from 2 to 4·5 MHz where terrestrial observation is not normally feasible. This range was chosen primarily as a compromise between the lowest frequency at which adequate sensitivity could be achieved with a loop antenna and the highest plasma frequency likely to occur for long periods. It was hoped that ionospheric focusing might help to improve the angular resolution on these occasions when the plasma frequency falls within the receiver band.[1]

The biggest single problem in low-frequency radio astronomy from space vehicles is the size of the antenna needed to achieve adequate sensitivity. The suitability of the *Ariel III* satellite for this purpose resides in the four booms which, when deployed, can support a large loop antenna enclosing an area of 4 m^2. A loop was chosen since no measurements with this type of antenna have yet been made and there is reason to believe that the bursts of noise recorded in earlier experiments with dipole antennas would either be absent or much reduced in observations of magnetic component of field. The low radiation resistance and high reactance of the loop antenna do however present some difficulty in covering a wide frequency band and the unusual features of the design all stem from this cause.

2. Receiver Specification

2.1 *Frequency Sweep*

In order to provide sufficient resolution for focusing effects to be observed, it was decided to use a band width of about 25 kHz and to sweep the receiver slowly across the frequency range measuring noise level with a conventional switching technique. The telemetry sampling rate available for the experiment is 1·15 per second and a sweep rate of 27 kHz per second was therefore chosen giving a total sweep time of rather less than 100 seconds.

In order to optimize the frequency resolution over the band and allow interpolation on the time axis, an accurately linear sweep is desirable and a maximum error of 1% in sweep rate was specified. Provision was made in the telemetered data for 8 measurements of spot frequency in each sweep.

2.2 *Sensitivity*

Calculation of the sensitivity required was based on an estimate of the mean free-space galactic flux and the relation between magnetic field strength and refractive index in the ionosphere.[2] In the region of interest where the extraordinary ray is approaching cut off the power flux is that of the ordinary ray and will be approximately one-half the free-space value. It was considered desirable that the accuracy with which this could be determined should be limited only by the digitizing interval of the telemetry and not by noise. For reasons discussed below it was decided that the gain should be adjusted so that this level of signal would produce an output of about one-third of the telemetry range. Since the digitizing interval of the telemetry system is 1% of maximum, the signal to noise expressed as a power ratio should be greater than 17 at this point when the response is linear. It is shown in the Appendix that the signal/noise ratio is worst at the low frequency end of the band and that a ratio of 17 should be achieved provided the time-constant of the receiver output is greater than 50 milliseconds. At the sampling rate of 1·15 per second there is therefore a useful margin in noise level

† Pye Telecommunications Ltd. (formerly Cambridge Works Ltd.), Haig Road, Cambridge.

while still making full use of the information capacity of the telemetry channel.

The margin is in fact somewhat greater than this would imply since a non-linear input/output relation was adopted to extend the range of measurement. Earlier satellite experiments with dipole antenna have shown bursts of high signal level in the region of the plasma frequency[3] and a response was therefore specified which would be linear at low levels and of decreasing slope for signals greater than the anticipated free-space value. A convenient means of achieving an output of this shape with a switched receiver is afforded by the use of an a.g.c. system in which the sum of the noise levels from the antenna and the dummy load is used for reference. The receiver then measures the ratio of the difference to the sum of these two. A compromise between accuracy and dynamic range was made, setting maximum output at 10 times the power level of the free space flux. Using the signal/noise ratios calculated in the Appendix, the predicted output is found to be one-third of maximum for the reference level of half the free-space flux.

In order to check the calibration of the receiver in flight, provision was made for the injection of a reference signal into the antenna circuit at a point near the low-frequency end of the sweep.

2.3 *Spurious Responses*

In view of the high sensitivity and wide sweep-range of the receiver it was felt that interference from ground transmissions might present a major problem. The most likely source was considered to be short wave stations in the range from 6 MHz upwards where the effective radiated power tends to be high and the ionosphere affords no protection. Fortunately, the main lobe in these cases is horizontally directed and the point along the orbit where maximum field strength occurs is at a considerable distance from the transmitter. Calculation of the field strength in typical cases showed that a rejection ratio of 70 dB for spurious responses above the receiver frequency should be adequate.

3. Principles of Operation

3.1 *The Signal Channel*

Figure 1 shows the general arrangement of the equipment with the main signal channel passing from left to right through the centre of the diagram. The homodyne principle is adopted since it simplifies the i.f. filters and has the special virtue in a swept receiver that the signal and local oscillator circuits track in synchronism. The central gap in the overall response due to the low-frequency cut-off of the i.f. amplifier has no particular significance when the wanted signal has the character of random noise.

The input to the r.f. amplifier is switched at a low speed between the antenna and a reference load. The detector following the i.f. amplifier is switched in synchronism and generates a square wave whose amplitude depends on the difference between the noise levels in the two conditions of the front end. The amplified square wave is measured in a further detector and after some processing feeds the telemetry encoder of the satellite. Ideally the reference load should be adjusted to equalize the noise levels in the absence of signal so that the output is zero in this condition. In practice it is difficult to avoid small changes in balance as the frequency sweeps across the band and to avoid the ambiguity of sign at low levels a small offset is introduced by feeding a fraction of the switching wave form into the last detector. The switching frequency of 35 Hz was chosen to be low compared with the lower end of the i.f. band but high enough to permit efficient smoothing by the time-constant following the final detector.

Figure 1 shows the 70 Hz pulse generator and frequency divider used to ensure an equal mark/space ratio in the switching cycle.

A calibration signal is injected continuously into the antenna circuit just prior to the switch. Since the receiver sweeps through its own band width for each data sample a c.w. signal of suitable amplitude would produce only two points on each sweep. To determine the amplitude and position of the calibration mark with sufficient accuracy, a signal spread over a range several times the band width is therefore required. Since the injection level is small, a noise generator provides a simple and convenient source and a filter defines the position and width of the mark. The mean current of the noise diode is monitored in flight as a check on the stability of the calibration signal.

3.2 *Antenna Tracking*

To maintain the required sensitivity it is desirable that the Q of the antenna should be as large as the bandwidth will allow and that the tuning should follow accurately as the receiver sweeps in frequency. This necessitates the use of automatic control for tracking the antenna circuit with the local oscillator since the loop is not a rigid structure and the impedance may be expected to exhibit anomalies in the vicinity of the plasma frequency. A tracking control circuit inevitably involves the presence of a locally-generated signal in the antenna to sense the resonance and this in turn leads to difficulties in antenna switching and in the suppression of spurious responses. These are discussed below. The gravest problem is, however, the interference caused by radiation of a signal sweeping across this band and inevitably accompanied by some proportion of harmonics. It was recognized at an early stage that this would create

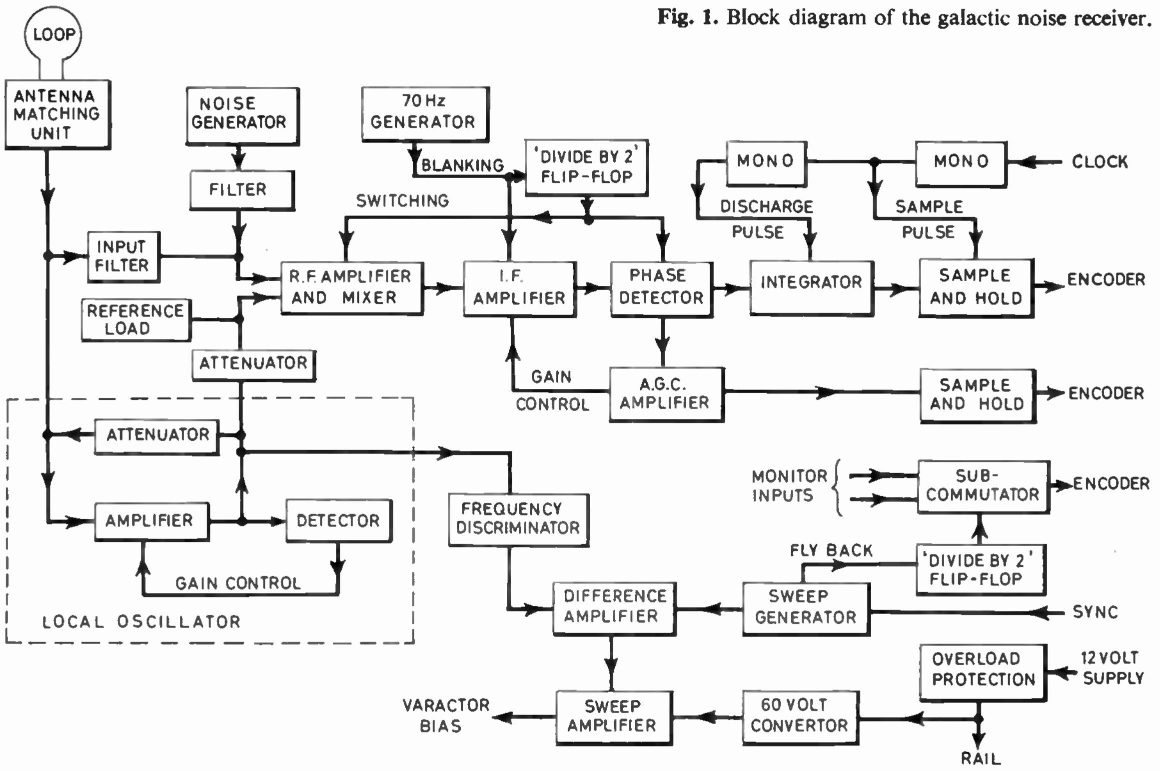


Fig. 1. Block diagram of the galactic noise receiver.

an intolerable situation for other experiments in the payload. The operation of the receiver is therefore restricted to one complete orbit in every five and it is switched off for the remaining four orbits.

Tracking is effected by coupling the antenna into the oscillator loop as shown in Fig. 1. Both circuits are tuned by varactor diodes supplied with the same bias voltage. The Q of the oscillator circuit is however much lower than that of the antenna so that the frequency is primarily determined by the latter. Compared with a system wholly dependent on the antenna resonance, this arrangement has the advantage that the receiver continues to sweep over the range with the antenna disconnected or coiled up in the stowed position. This enables most of the receiver to be checked with the satellite in the pre-launch condition.

The oscillator level is stabilized with a control loop for the following reasons:

(*a*) The contribution of noise from the local oscillator is reduced by extending the response of the control circuit above the i.f. band.

(*b*) Since the reference load is also fed from the local oscillator for reasons described below, signals from the antenna can reach the receiver in the reference condition via the amplifier of the oscillator circuit. This effect is reduced by the loop gain of the level control.

(*c*) The frequency measuring circuit which forms part of the sweep control system requires a constant level for correct operation.

The state of the local oscillator a.g.c. circuit provides a measure of the resonant impedance of the antenna. The behaviour of a loop antenna immersed in the ionosphere is of some interest apart from the main objective of the experiment and provision is therefore made for monitoring the local oscillator a.g.c. in flight.

The use of antenna tracking with a switched receiver presents a major problem in the i.f. amplifier. The local oscillator signal appearing in the antenna is chopped by the antenna switch and appears at the mixer in bursts at a rate of 35 Hz. The effect is minimized by feeding a comparable signal to the reference load but the levels cannot be made to balance precisely, and the mixer output contains a 35 Hz square-wave which is generally about 80 dB above the signal which it is desired to measure. In view of the high harmonic content of this square-wave falling within the i.f. band, a blanking stage is introduced which inhibits the amplifier for the duration of each transient. To be effective it is necessary to exercise care in controlling the low-frequency response prior to blanking. Figure 1 shows the blanking wave derived directly from the 70 Hz oscillator. The small inherent delay in the divider ensures that blanking starts before the switching transient.

3.3 *Data Outputs*

The operation of the satellite encoder requires the signals presented to it to remain constant over the encoding period.[4] To make full use of the data rate, the main output of the receiver is therefore integrated over the period of each sampling interval and the integrand transferred to a holding circuit. The integrating capacitor is reset at the start of each interval so that successive samples are independent from this point. The two monostable circuits shown in Fig. 1 produce the delays required for the timing of the sequence in relation to the clock pulse supplied from the satellite encoder.

In addition to the main output from the receiver the a.g.c. is monitored by eight samples in each sweep period. This is useful for two reasons: a comparison of the main output with the a.g.c. level is helpful in identifying interference, and the a.g.c. calibration can be used to extend the range of measurement of signals which overload the main channel. Since the a.g.c. level can change at about the same rate as the signal, a sample and hold circuit is also used in this case.

A number of other parameters are also monitored by less frequent samples in the telemetry format to check the operation of the equipment in flight. These are frequency, varactor bias, local oscillator a.g.c. and the current in the noise diode which generates the calibration signal. The decision to monitor antenna behaviour from the local oscillator a.g.c. was made at a late stage in the development programme when the allocation of the telemetry channels had already been made. Since very infrequent samples of calibration diode current would suffice, this channel was changed for the local oscillator control and the calibration diode substituted for half the frequency readings on alternate sweeps. The commutator shown in the block diagram (Fig. 1) is switched from a divider triggered by the fly-back of the frequency sweep.

3.4 *Frequency Sweep*

In view of the non-linear characteristic of the varactors the sweep is controlled by reference to a frequency measuring circuit with a linear response. A ramp generator produces a linear voltage/time output which is compared with the discriminator in a differential amplifier and the output of this supplies bias to the varactors. The frequency then follows the linear ramp with an accuracy determined by the gain of the differential amplifier and the slope of the varactor characteristic. Since the latter varies widely across the control range, the differential amplifier includes a two-step diode function generator to minimize the changes in open loop gain.

The sweep period was chosen to be 112 seconds or four periods of the slowest pulse available from the spacecraft encoder. The forward sweep normally occupies about 96 seconds, fly-back 4 seconds and the sweep then waits at maximum frequency for the next synchronizing pulse when the cycle starts again. The wait period is designed to cover telemetry of several parameters which can then be related in the static condition. In the absence of a synchronizing pulse the sweep will run free with a somewhat longer period.

4. Interference Rejection

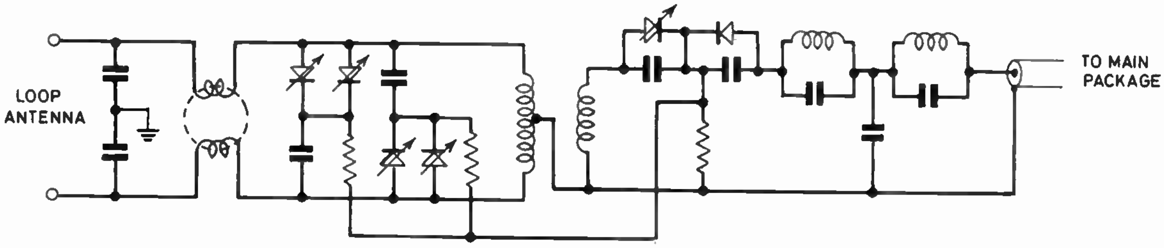
A low-pass filter is placed between the antenna and the r.f. amplifier to minimize interference at frequencies above the receiver band. This filter is shown in Fig. 2 immediately preceding the r.f. amplifier. Since the receiver covers a band of more than two to one in frequency, the second harmonic response cannot be excluded if the filter is fixed. It was therefore decided to incorporate varactor diodes in the filter so that the cut-off could be varied with the frequency sweep. The range of cut-off frequency had to be restricted to avoid a too-violent change in filter impedance. Nevertheless a substantial improvement in the rejection of spurious signals was obtained by this means.

As described above, the presence of local oscillator signal at the antenna cannot be avoided, and since the varactors are essentially non-linear devices, particularly at low values of reverse bias, cross-modulation can produce in band signals at this point and subsequent filtering is ineffective. The only discrimination preceding the varactors is the resonance of the antenna itself. Substantial reduction in the response to harmonics was obtained by reducing the level of the sensing signal as far as possible and by arranging the varactors in symmetrical pairs so that even-order harmonics are balanced. The target rejection ratio of 70 dB for spurious signals was achieved for most of the range but falls to 65 dB at the low end of the band.

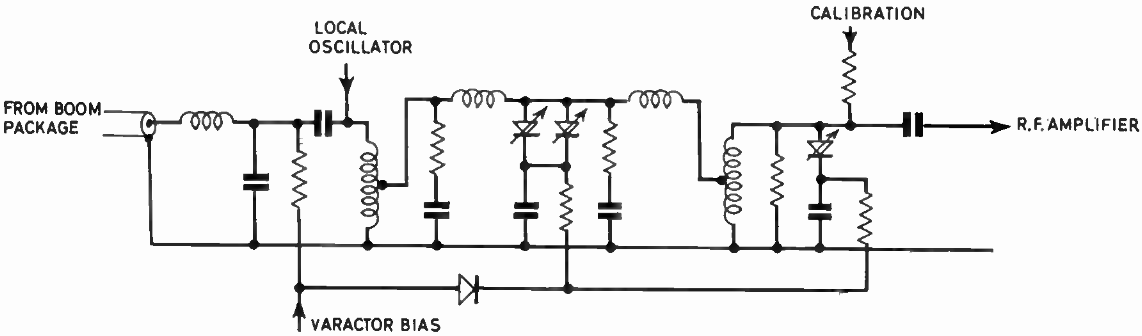
It was found necessary to introduce additional circuits in the antenna matching unit to deal with interference from two sources in the satellite itself. These were the telemetry transmitter at 136 MHz and a signal at 29 MHz applied to the ionospheric probe in the Birmingham University experiment.

The telemetry antennas are located at the opposite ends of the vehicle and are symmetrical with respect to the loop antenna. Nevertheless, during integration testing the telemetry transmitter was found to have a significant effect on receiver output and a level of 350 mV at 136 MHz could be measured in certain conditions between the ends of the loop antenna and the boom.

An effective solution was found by connecting the antenna through bifilar windings on a small inductor



Boom Package.



Main Package.

Fig. 2. Circuit diagram of the front end, showing connections between the antenna, the boom unit and the main package.

which isolates the receiver for unbalance signals at v.h.f. (Fig. 2).

The two booms at right angles to the one used for connections to the loop antenna both carry ionospheric probes and one of these is excited with a substantial signal at 29 MHz.[5] One side of the antenna passes within a few inches of this probe and an amplitude of about 1 V was measured at the antenna terminals from this source. The effect on the receiver is to raise a series of peaks as the frequency sweeps through sub-multiples of 29 MHz. It was found that the interference could be effectively removed by the addition of resonant sections at the output of the antenna matching unit and these were incorporated in the flight models.

5. Construction

The connections to the loop antenna are made at the end of one of the four supporting booms and the receiver is located inside the satellite shell at a distance involving about six feet of connecting cable. It was found necessary for the tuning varactors, balance transformer and some filters to be connected directly to the antenna and these are housed in a small box at the end of the boom. The exposed position and small thermal capacity of this box give rise to wider variations in temperature than those of the main package. A surface treatment was devised by Semple[6] to raise the mean temperature as far as possible since

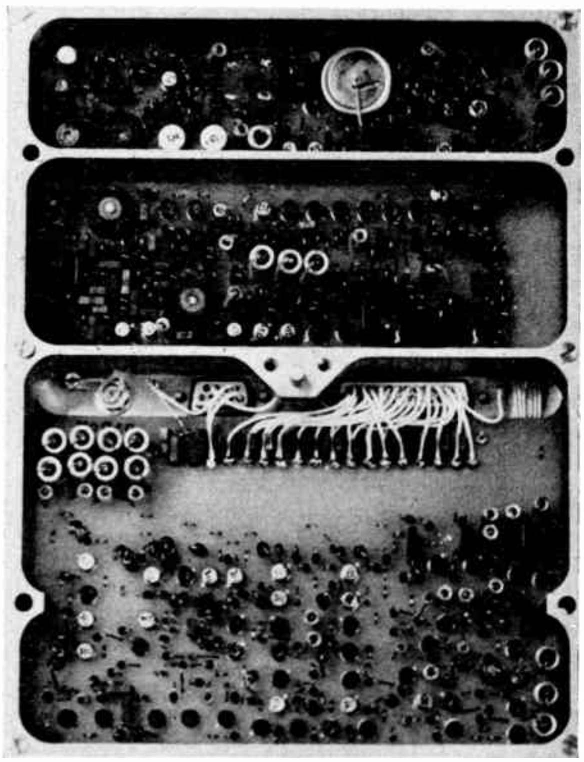


Fig. 3. View of the receiver, before foam encapsulation, showing (top to bottom) the low-frequency and data handling section, the r.f. and i.f. section and the frequency-sweep and converter section.

the greatest change in performance was found to occur at the low temperature extreme.

With the exception of the antenna and the boom unit all parts of the receiver are housed in a flat package of simple design bolting directly to the cruciform structure in the satellite. Figure 3 shows a view of the equipment before foam encapsulation and with the cover plate removed. The housing is machined in one piece and has three compartments to provide screening for r.f. and i.f. circuits and isolation of the converter. The components are mounted on cards bonded to the frame. The weights of the main package and the boom unit are 1·7 kg, and 90g respectively and the total power consumption is 320 mW.

6. Preliminary Results

The frequency sweep and the other telemetered performance parameters indicate that the receiver is operating normally in flight and that the antenna deployed correctly. However, the main output and the a.g.c. channel both show strong signals changing in synchronism with the probe switching of the University of Birmingham experiment.[5] The interference is believed to arise from modulation of the local plasma by the Birmingham probe since the effect was absent in ground measurements and shows a marked correlation with the measured electron density. A study of possible mechanisms is in progress. In certain conditions of the electron temperature mode the interference is small enough to be neglected and useful measurements of cosmic noise have been made in this case.

7. Acknowledgments

The work described was carried out on behalf of the Nuffield Radio Astronomy Laboratory in the University of Manchester at Jodrell Bank. Throughout the project, contributions to the design were made by many members of the Laboratory under the direction of Professor F. G. Smith.

The antenna and deployment mechanism were designed by the British Aircraft Corporation who also carried out the integration tests.

The author is indebted to many of his colleagues who helped with the project and to the Directors of Pye Telecommunications Ltd., for permission to publish this paper.

8. References

1. Smith, F. G., 'The possibility of observing features of galactic radio emission from a satellite borne radio telescope,' *Mon. Not. Roy. Astronom. Soc.*, **122**, p. 527, 1961.
2. Hugill, J., 'The measurement of random r.f. field by the use of short electric and magnetic dipoles'. *Planet. Space Science*, **8**, No. 2, pp. 68–70, 1961.
3. Hugill, J. and Smith, F. G., 'Cosmic radio noise measurements from satellite *Ariel II*,' *Mon. Not. Roy. Astronom. Soc.*, **131**, No. 2, pp. 137–53, 1965.
4. Laws, A., 'The data handling system for the *Ariel III* satellite', *The Radio and Electronic Engineer*, **35**, No. 2, pp. 83–7, February 1968.
5. Wager, J. H., 'The University of Birmingham electron-density and temperature experiment on the *Ariel III* satellite', *The Radio and Electronic Engineer*, **35**, No. 1, pp. 55–63, January 1968.
6. Semple, E. C., 'Thermal design and thermal testing of the *UK-3* spacecraft', *The Radio and Electronic Engineer*, **35**, No. 1, pp. 41–8, January 1968.

9. Appendix

It may be shown that the power density available at the receiver from a single-turn loop is given by

$$P = 2\pi^2 A^2 ZQfI/3c^2 L \quad \text{watts/Hz.}$$

where A = antenna area (m^2)

Z = space impedance (Ω)

Q = loaded Q of the antenna

f = frequency (Hz)

I = intrinsic brightness W/Hz/m^2/sterad

c = velocity of light (m/s)

L = antenna inductance (H)

The area and inductance are fixed by the geometry of the antenna and the only factor at the disposal of the designer is Q. In practice, this is restricted by the band width and the signal is therefore smallest at the low-frequency end of the band. To assess the power available in this case we put $Q = f/B$ where B is the band width in Hz.

When I has half the estimated free-space value we write

$$A = 4\ \text{m}^2$$
$$L = 10\ \mu\text{H}$$
$$I = 8\times 10^{-21}\ \text{W/Hz/m}^2\text{/sterad}$$
$$f = 2\times 10^6\ \text{Hz}$$
$$B = 25\times 10^3\ \text{Hz}$$

and obtain $P = 5{\cdot}64\times 10^{-20}$ W/Hz ......(1)

Since the radiation resistance is small compared with the antenna losses, the total excess noise from antenna and receiver is

$$P_E = KT_1 + KT_0(F-1) \quad \text{watts/Hz}$$

where K = Boltzmann's constant

T_1 = antenna temperature

F = noise factor

T_0 = reference temperature at which F is defined

When $T_1 = T_0 = 300°K$

$F = 3$

$P_E = 1{\cdot}23\times 10^{-20}$ W/Hz ......(2)

For a switched receiver the mean square noise power at the output is given by

$$\Delta P = \sqrt{2}\,(P_1^2+P_2^2)^{\frac{1}{2}}/(BT)^{\frac{1}{2}} \qquad \ldots\ldots(3)$$

where P_1 = power level with the receiver connected to the antenna

P_2 = power level when connected to the reference

t = output time-constant

The receiver measures the difference between the signals from the antenna and the reference. For a linear system the signal/noise ratio is therefore

$$S/N = (P_1^{\frac{1}{2}}-P_2^{\frac{1}{2}})^2/\Delta P$$

Substituting for ΔP from (3)

$$S/N = \sqrt{Bt}(P_1^{\frac{1}{2}}-P_2^{\frac{1}{2}})^2/\sqrt{2}(P_1^2+P_2^2)^{\frac{1}{2}} \qquad \ldots\ldots(4)$$

Since P_1 and P_2 are equal when there is no signal

$$P_1/P_2 = (P+P_E)/P_E$$

Substituting numerical values from (1) and (2), equation (4) becomes

$$S/N = 0{\cdot}48\sqrt{Bt}$$

To achieve a signal/noise ratio of 17 the output time-constant t, should be 50 ms.

Manuscript first received by the Institution on 30th November 1967 and in revised form on 20th June 1968. (Paper No. 1222/AMMS 19).

Conferences in 1969

Marine and Shipping Conference

An International Marine and Shipping Conference is being organized by the Institute of Marine Engineers in collaboration with the Royal Institution of Naval Architects. The Conference will be held in London from 10th to 20th June, 1969, and some 80 papers have been accepted from authors throughout the world.

Subjects to be covered include: ship operation and management; main propulsion machinery; ancillary shipboard services; cargo handling and stowage; marine safety; dry docking and repairs; miscellaneous craft; ocean engineering; stern gear, shafting and propellers; materials; education, training and manning; electrical and control engineering; ship types of the future, ship design and construction. There will also be visits to industrial organizations.

The basic registration fee will be £10, and further details may be obtained from the IMAS 69 Conference Office, Institute of Marine Engineers, 76 Mark Lane, London, E.C.3.

Lasers in Medicine

A Conference on Lasers in Medicine is being organized by the Institute of Physics and The Physical Society in collaboration with the Royal Society of Medicine, the Ophthalmological Society of the United Kingdom and the Hospital Physicists Association. Professor O. S. Heavens (of the University of York) will be the chairman of the Conference, which will be held in the Medical School of the Middlesex Hospital, London, from 2nd to 4th July 1969.

The conference is intended to interest clinicians and medical research workers, and physicists and engineers working in the application of these techniques in medicine. Original contributions are invited under the following subject headings:

Biological and physical effects of laser irradiation; clinical applications; laser engineering for medical applications; instrumentation and dosimetry; hazards and safety aspects.

Offers of papers should be sent with 300-word synopses to the Conference Scientific Secretary, Mr. R. Oliver, Department of Radiation Physics, Churchill Hospital, Oxford, by 14th February 1969.

Advance registration will be necessary. Further information on the Conference is available from the Meetings Officer, Institute of Physics and The Physical Society, 47 Belgrave Square, London, S.W.1.

Symposium on Space Technology and Science

The Eighth International Symposium on Space Technology and Science will be held at the Nippon Toshi Center in Tokyo from 25th to 30th August, 1969.

Papers are invited on topics dealing with all related subjects, including propulsion; materials and structures; flight dynamics, astrodynamics, aerodynamics; environment; spacecraft and rockets; space electronics; guidance and control systems; engineering; space science (payload design, instrumentation); medicine and biology; national space programmes; space law.

Synopses of papers should be submitted before 30th April to Professor Jiro Kondo, 8th ISTS-Tokyo 1969, Department of Aeronautics, Faculty of Engineering, University of Tokyo, Bunkuo-ku, Tokyo 113.

Registration forms and further information on the Symposium are available from the General Secretariat, 8th ISTS-Tokyo 1969, Japanese Rocket Society, c/o The Yomiuri Newspaper Building, 1-2, 3 chome, Ginza, Chuo-ku, Tokyo 104, Japan.

Radio Engineering Overseas . . .

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

INDUCTORLESS FEEDBACK INTEGRATED CIRCUITS

Transformers and inductors have played a very important part in the construction of analogue amplifiers up to the present day. In integrated amplifiers, however, these elements will have to be largely foregone since only very small inductance values can be realized, for instance, in thin-film technique. A combination of integrated circuitry and conventional inductors does not seem satisfactory since essential advantages of the integrated circuits are not fully realized.

The introduction of a new technique cannot be welcomed if important potentialities of earlier techniques must be sacrificed. For instance, with the aid of transformers, it is possible, by means of combined series-parallel feedback, to match amplifiers to source- and load-impedances only with negligible losses. In the case of flat as well as frequency-dependent amplification, which is determined by equalizer networks in the feedback path, the input and output impedances in the transmission range remain flat. This problem is now solved without the aid of transformers. A paper based on work carried out at the Central Communications Laboratories of Siemens A.G., Munich, describes the circuits devised to offset some of the drawbacks which result when transformers cannot be used. By active matching of the amplifier input according to the methods outlined here, noise figures have been obtained, which are, for example, as low as those for corresponding amplifiers using transformers.

'Feedback circuits for integrated amplifiers' (in English), H. Gölz, *Nachrichtentechnische Zeitschrift*, **21**, No. 7, pp.374–7, July 1968.

IMPROVED CHANNEL-CAPACITY OF SATELLITE COMMUNICATION SYSTEMS

For multi-station operation of a satellite communication system with frequency multiplexing of the ground station signals and frequency modulation (with reference to super-group multiplexing in the relay trunk at radio frequency), the telephone channel distribution between the ground stations has maximum flexibility when each call is transmitted by an individual carrier. This type of system is not optimal from the power point of view, since fairly high peak power is required from the satellite transmitter.

The efficiency can be considerably improved if use is made of the call statistical properties. A conversation is well known to occupy a channel discontinuously: there are pauses between words, meaningful pauses between sentences, and finally, each subscriber is quiet for half the time on the average while listening to his interlocutor. If the ground station signals can be disconnected during conversation pauses, the relay trunk occupation time, and hence the trunk average power loading, can be reduced. This in turn enables either the relay transmitter power to be reduced, or the system capacity to be increased by loading the relay with extra signals.

The pauses in conversations are widely utilized for increasing capacity in communication systems. For instance, there are systems in which the pauses are used for data transmission, i.e., each telephone channel is combined with a data transmission channel.

In a Soviet paper a multi-station satellite system is described in which the telephone calls are transmitted on individual f.m. carriers and the capacity is improved by suppressing the carriers during modulation pauses. The statistical properties of the multi-station signal are discussed. The suppression effect is estimated in relation to power, and the problems arising in such suppression are examined.

'Improving the channel capacity of satellite communication systems with multi-station access and frequency multiplexing', V. L. Bykov, V. A. Borovkov and S. M. Khomutov, *Telecommunications and Radio Engineering* (English language edition of *Elektrosvyaz* and *Radiotekhnika*)., No. 1, pp. 1–5, January 1968.

VOLTAGE CONTROLLED MULTIPLIERS WITH PUNCH-THROUGH DIODES

Recently punch-through diodes have begun to be widely used because of their high efficiency and linear input versus output characteristics when they are used in charge-controlled-mode frequency-multiplier circuits. When these diodes are used in voltage-controlled mode, voltage in sharp spike shapes is generated and a higher harmonic multiplication is obtained.

Specially designed step recovery diodes of p-i-n construction were used in the past in voltage-controlled multipliers, and it was explained theoretically that the pulsating current caused by the disappearance of the storage charge was the source of the harmonic generation. However, in reality, the Q in diodes has a limited value, and the voltage wave shapes between the series resistance in diodes are considered to cause high multiplication.

A paper based on research carried out at Fujitsu Ltd. in Japan, analyses cases where Q has a limited value and furnishes material for the design of a voltage-controlled multiplier using punch-through diodes.

'An analysis of voltage-controlled multipliers with punch-through diodes', Tatsuo Miyakawa, *Fujitsu Scientific and Technical Journal*, **3**, No. 2, pp. 43–67, September, 1967.