The Journal of the

BRITISH INSTITUTION OF RADIO ENGINEERS

FOUNDED 1925 INCORPORATED 1932

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

VOLUME 21 JANUARY 1961 NUMBER 1

FOREWORD

I AM very glad to have the opportunity of writing this foreword to the first number of our *Journal* in its new form. With all the advances that have been made in radio and television and their recording, no substitute has yet been found, as a basis for study and reference, for the written word; and I am sure that all my fellow members will welcome the Council's decision to devote a greater part of our resources, in future, to the publication of more papers and technical information than ever before.

Within the thirty-six years of the Institution's existence scientific knowledge and application have developed more rapidly than at any previous time in man's history; and I think that as an Institution we can be justified in feeling that we have risen to the challenge of these four momentous decades. This is not the place for a catalogue of achievements, but it is well that we should remember the vital part that the Institution has played, since the early days of wireless, in securing facilities for education and training in radio and electronic engineering; in the exchange of information about our science and profession; and in securing the establishment of standards in engineering performance—an aspect, incidentally, which was not tackled until 1925.

But whilst our members, especially among the rising generations, must never lose sight of the aims that have guided the prodigious growth of our Institution, whose contribution is now internationally recognized, our main responsibilities in the years ahead must be to keep abreast of the tempo of scientific progress, which, especially in our own field, continues to grow at an exponential rate. This means, in particular, that we must have no reluctance in publishing quickly, and in discussing freely at our meetings and conventions, any paper of a provisional nature whose potential value would benefit those who might be able to make immediate use of it and thus help to keep us ahead in the development of our science.

We are on the threshold of space exploration in which Electronics will of course play a vital part; and this aspect will be the main feature of the Institution's 1961 Convention. It is in the belief that the past achievements of the Brit.I.R.E. justify my confidence in its future that I welcome this opportunity to send my good wishes to every member of the Institution for the coming year.

Mountbatten of Burma A.F.

Vice-Patron
British Institution of Radio Engineers

INSTITUTION NOTICES

New Year Honours

The Council of the Institution has congratulated Wing-Commander William Albert Russell, B.Sc., R.A.F., on his appointment as an Ordinary Member of the Military Division of the Most Excellent Order of the British Empire. Wing-Commander Russell has been in the Education Branch of the R.A.F. since 1939; from 1957 to 1960 he was at R.A.F. Locking and he has recently been appointed senior education officer at R.A.F. Melksham. He was elected an Associate of the Institution in 1950 and was transferred to Associate Membership in 1951.

Indian Recognition of the Institution

Advice has just been received that the Ministry of Education of the Government of India recognizes Graduate membership of the Institution as a qualification equivalent to an engineering degree in telecommunication for the purpose of recruitment to higher posts in Government service. The only stipulation is that candidates who obtained Graduateship before November 1959 will be required to take or obtain exemption from the papers on "Mathematics" and "Principles of Radio and Electronics".

This ruling places Indian Graduate members of the Institution on an equal footing with those who hold similar qualifications from other engineering institutions or engineering degrees. Similar recognition has already been given in the United Kingdom and elsewhere in the Commonwealth.

Symposium on Tunnel Diodes

The joint meeting with the Electronics Group of the Institute of Physics and the Physical Society will be held on 7th February at Caxton Hall, London, S.W.1. Members who have applied for and received tickets are asked particularly to note the changed venue.

Despite the move to a large hall, the meeting has been heavily oversubscribed and it is regretted that *no further tickets are available*. It is expected that most of the papers will be published in the *Journal* of the Brit.I.R.E. or the appropriate publications of the Institute of Physics or in both.

Binding of Journals

The index for the 1960 Volume of the *Journal* was circulated with all copies of the December 1960 issue. Members wishing to have their *Journals* bound by the Institution should return their copies together with the index and a remittance for 16s. 6d. The appropriate return postage (3s. for Great Britain, 4s. Commonwealth and other countries) should also be included.

Students' Essay Competition

On the recommendation of the Education Committee the Council has agreed to widen the conditions of entry, increase the value and give a wider choice of essay subject for the Institution's student award. The next competition will be held during the 1961–62 session. A choice of subjects will be announced in October, and closing date for receiving entries is 28th February 1962.

The competition will be open to any *bona fide* student of Radio or Electronic Engineering attending full-time or part-time courses of study; the maximum age will be 26 years and the prize will be valued at £15 15s. Although suggested subjects will be published, candidates may, if they wish, write instead a survey paper on any subject coming within the fields of the Institution's activities. In this event it is recommended that candidates should seek the advice of the Examinations Committee on their choice of topic.

J.Brit.I.R.E. Volumes

As stated in the 34th Annual Report of the Council† an increasing number of papers are being accepted for publication in the Institution's *Journal*. Although the new format will permit of some increase in editorial content per page, the new overall size of the *Journal* will be too large and cumbersome to warrant an annual volume.

For ease of reference, therefore, the present issue is the first of six issues which will comprise Volume 21 of the *J.Brit.I.R.E.* Commencing from the July 1961 issue, a new volume (Vol. 22) will be started and thereafter six monthly *Journals* will comprise a new and separate volume.

List of Members 1961

The ninth edition of the Institution's "List of Members" will be published in March. Copies will be sent to all Corporate Members, Associates and Graduates whose names appear therein. Registered students are *not* included, but they may purchase copies of the List, price 5s.

The membership lists have been corrected to 31st October 1960. The publication also includes details of the membership of all the standing Committees.

Correction

The following amendment should be made to the paper "Electron transmission of mesh lenses for scan magnification in television picture tubes" which was published in the December *Journal*.

Page 916: the expression in the line before eqn. (18) should read

$$E_2 = -E_1 \equiv E.$$

† *J.Brit.I.R.E.*, December 1960.

U.D.C. 534.7

Electro-Acoustics for Human Listeners

Address at the Inaugural Meeting of the Electro-Acoustics Group, held in London on 12th October 1960

By

Professor COLIN CHERRY, D.Sc.†

Summary: The way in which human listeners distinguish sounds is discussed and its applications in speech compression systems considered. The relationship between speech perception and speech production is dealt with. Stereophonic hearing is shown to depend on the separation of images and judgment curves for binaural separability are presented. These have led to the construction of a mathematical model for their computation. The present theories of stereophonic hearing are however shown to be incomplete and a plea made for further study. Several other lines in electro-acoustics calling for research are indicated. A better appreciation of the role of the listener would lead to improved communication systems.

What is it that an electro-acoustic transducer, such as a microphone, a loudspeaker or a stereophony system, is really required to do? Traditionally it is to convert the sounds uttered by human beings or musical instruments from one energy-form to another, but, and this is the point, to do this *as accurately as possible.*

I want to start then by what may seem to many of you to be a heresy—by denying that this is the real, ultimate, aim of our systems. But a little heresy may not be out of place in an inaugural address, when we are essentially looking to the future; for as Dr. Johnson said so pointedly: "all great truths begin as heresies". This address is intended as a Birth Certificate, not an Obituary Notice, so I will not adhere to the respectability of tradition and past glories, but put down some facts about our future life . . . as I see them.

The faithful reproduction of speech and musical sounds has sufficed in the past, with good reason, and will indeed serve as an ultimate aim for a long time to come, if we are thinking only of the simple conditions of broadcast and gramophone listening. It is my purpose today to stress that it is not enough however. In the last analysis it is a false criterion. But I have not come here to tell competent designers of high-fidelity equipment how to design their systems: all I want to do is to focus attention on the least understood links of the whole telecommunication chain—the human speaker and listener. Our knowledge of acoustic science and our principles of design have reached the most sophisticated levels; in comparison, our understanding of what goes on inside the human terminal is pitifully small. We need to know much more; that is all I want to say, and to give my own view that all *major* future progress will depend upon such better understanding, not solely upon more research into physical acoustics.

Now I know full well that any designer of loudspeakers realizes the need to test his results on a live audience, listening in typical furnished rooms, or in artificial noise simulating aircraft cabins, tanks and other difficult situations. After making all his physical measurements he relies on listening tests, with trained or untrained listeners—tests often called "subjective" and therefore, in some physicists' minds, not quite respectable, dodging the question they say . . . "not scientific". Actually, this view shows a misunderstanding of the term "subjective", because there is no reason why tests involving human beings should not be made objective—if good method and the techniques of experimental psychology be used—but I will return to this later.

I have at last mentioned the magic words, the Open Sesame . . . "experimental psychology". Not descriptive or clinical psychology, of course, but strictly *experimental*; that is, we apply stimuli to people and watch how they jump. Strict behaviourism is, however, not enough. As in all science we not only watch what happens, but we set up theories and test these by further experiments which they suggest. We build theoretical models, often mathematical models, and predict as yet unknown phenomena with these. Can we make models of psychological processes? Can we used mathematics and predict new results? Indeed we can, and I want to give you some examples later.

But first, I should like to give a very simple illustration of the difference between a psychological and a physical phenomenon. Listen for a moment to a record: of a conversation with someone hammering in the background.

† Imperial College, London, S.W.7.

[A record was played at this point, of a conversation half-drowned by the noise of hammering.]

You can focus your attention on the conversation here if you wish—identifying almost every word. Now listen again and try to *count* the number of blows of the hammer. . . . I would judge that this time you did not hear what the speakers said. I doubt whether you noticed that this time they spoke in Hungarian!

Acoustically speaking, in the air itself, the sounds of speech and hammer blows are inextricably mixed up. No amount of instrumental measurements will separate them. As soon as they enter your head however, they are separated into separate images, and you can respond to either separately. By what remarkable filtering process does the brain separate them? At present we cannot say, yet it would seem to me that this problem is the most fundamental in the whole of telecommunication. The aim of our systems is to create these images in the listener's mind, to keep them distinct and separate as required. Yet if we do not understand the process, how can we control it?

The traditional view that waveform fidelity is the ultimate criterion ignores this process of image-formation, because it aims not only to reproduce the images of speakers, or of musical instruments, faithfully, but of the background noises as well.

These separate and distinct images are called by psychologists by the German word *gestalten*. All the mass of sense data which falls upon our ears, our eyes, our touch does not remain an incoherent mess; it congeals into separate *gestalt* images we call the *things* that make up our world, towards which we can selectively respond. I see a chair and can sit on it, or a hat and can put it on my head. I hear my name called and answer to it.

A new-born baby cannot do this—sense data still form a confusing mess; at first he cannot tell the inside of a cup from its outside; he cannot separate one voice from another. Only by the most laborious and lengthy learning processes do the *gestalten* eventually form; the child can then separate the various demands that come to his attention; he can select and adapt to the world.

The word *learning* is the key. The first great difference between a human being and an inanimate physical object is memory. The human has this great store of his past experience so that, when he receives a stimulus such as the sounds of someone's voice his response depends not only upon the stimulus itself but upon his whole past experience as well.

In the case of speech the store in each one of us is specially vast, and of the greatest importance to us. Although in the course of our lives we listen to many thousands of voices, all different in accent, speed, stress and pitch, nevertheless we each build our store which represents our response habits. Clearly we cannot store the whole mass of sound-waves which we have ever heard, but our brain must abstract certain average *parameters*. That is, what we would call a statistical store, which is modified slightly every time we hear someone speak.

What are these parameters, in detail? We do not know precisely, but we can say a lot about them. For one thing we do not use exactly the same ones every time we make a perception but, being organisms, we can adapt and can operate differently according to circumstances (for example the parameters we use may depend upon the noise present). Now voices, music and all the other sound sources we listen to, provide our ears with an enormous mass of data, and the first thing the brain does is to throw most of it away; first, peripherally at the cochlea and then at various stages of the inner and central nervous system, until by successive abstractions and transformations sufficient parameters remain for recognition, and image formation, to take place. The analysis performed and the parameters which are identified will, as I have said, most certainly vary from time to time, but we may think of the whole process as *goal-seeking*—as a continual groping, changing the mode of analysis so as to maximize our chances of success in various circumstances of noise, disturbance and environment.

Various telecommunication systems have been built, of which the Vocoder is perhaps the best known, which seek to short-circuit this brain abstraction process by themselves abstracting what are thought to be the acoustic parameters of speech recognition, and presenting these compressed sounds to your ears. They work remarkably well, but they operate in a fixed way, not adapting their behaviour to the listener's needs. If we knew more about this human adaptation process, in detail, perhaps we could improve our speech compression systems by changing the data they extract from the speaker's voice, or by processing it in various ways, according to the noise or other factors.

One of these speech compression schemes, that originated by Walter Lawrence at S.R.D.E., Christchurch,† is of particular psychological interest here. Briefly, for those of you who may not know this system, the machine at the transmitting end measures only half-a-dozen parameters of the speaker's voice automatically—the varying pitch and intensity of his larynx tone and the fluctuations of the main vocal cavity resonances (the formants). These greatly reduced and slowly-varying data are sent and, at the receiver end, control a set of oscillators to synthesize speech.

[A record was played at this point.]

† Signals Research and Development Establishment, Ministry of Aviation.

The remarkable success of this system[1] using, it must be remembered, only six parameters to describe the motions of the speaker's vocal organs, underlines a fundamental point about our perception of speech, which is this: that our perception of speech is undoubtedly reinforced by our own vocal habits. For we not only hear speech, but produce it. Speech, in this sense, forms a class of sounds totally distinct from all other sounds of this world. When we hear the sounds of someone's voice we can associate them with the movements of our own vocal tract.[2] It seems to me logical then that the parameters of speech recognition might be identical with the parameters which efficiently describe its production—larynx pitch, formants and perhaps others.

In my department at Imperial College we have made a number of experiments a few years ago which also stress the close relation between our speech perceptual habits and our speech production habits. One simple one we called the technique of *shadowing* which I can illustrate here.[2] A speaker sits before a microphone reading steadily from a book—any book written in common English—whilst a listener, wearing headphones, listens and repeats concurrently. He is then "tracking on to" or "shadowing" the speaker. The remarkable thing is that he makes no errors.

[A record illustrating shadowing was then played.]

How easily the listener's speech habits are called upon! If now we use an American speaker and an English listener their speech habits are slightly different—but the listener responds by translating; that is, to his own *speech* habits.[2] We are slaves to such habits; even stammerers cease to stammer when "shadowing" normal people.[3]

Although we all have different accents, have different interests and all "speak our minds" we are none the less conformists to the statistical laws of our language.[4] However hard we try to be original, we unknowingly adhere to these laws, and speech communication works. These laws, our habits, prior probabilities, whatever you choose to call them, dominate what we say, and what we hear too.

Now to return to my original theme: the separation of sound images or *gestalten*. How do these speech habits, these prior probabilities, enable us to separate a voice from its noisy background? Since the noise is a chance process there is only one basis of action open to the brain; namely, statistical inference. All the brain can do is to form a set of alternative hypotheses (weighted according to past experience, prior probabilities, habits) and test these against the sound evidence reaching the ears; this testing procedure can follow a number of courses which I cannot describe here and now. But the mathematical ways of describing these various procedures of statistical inference are known and are collected together in the general field called "Statistical Decision Theory".[5] There is little doubt that such theory will aid us to understand aural and other modes of perception.

Rather than outline this mathematics, I will now describe some more experiments which may serve instead to illustrate the concepts.

But first let me point out one fact about human beings, which so far seems to have evaded the notice of most communication engineers—namely, that we have two ears. Most people imagine that the possession of two ears gives, as its main consequence, the faculty of directional hearing. But really this is secondary. The primary faculty is not this, but is a greatly enhanced ability to separate a voice from noise, or one voice from another as at a cocktail party. The sounds reaching the two ears provide slightly different sets of evidence, by which the inference processes can be carried out more efficiently. So far, no telecommunication system has taken real advantage of this faculty—not even the so-called stereophony systems.

Directional hearing—this is a *consequence* of better image separation, for the subjective sensation we call "space" is nothing but the distribution of images in it. As a consequence of enhanced image separation we say we hear one voice over *here* and another over *there*.

Real understanding of stereophonic hearing then hangs first on understanding of image separation; this in turn depends upon something even more basic, namely how does the binaural fusion process take place, by which the brain accepts the different data at each ear and relates them so that we hear only one world? Believing this to be the fundamental question we have concentrated first on it, in my department, and I should like to illustrate my arguments with a few specimen results.

If you listen to sounds through headphones, directional hearing is lost, because head-turning now has no effect; however, better image separation still exists. Listening to one source, say a voice, this appears to have a certain "size" and "position" in the subjective space inside and slightly to the back of your head. "Subjective" in the sense of your private sensations, but we can measure the "size" and "position" objectively, as I mentioned at the beginning of this talk. The method is a typical one of experimental psychology.[4, 6]

To do this we first argue that the finite "size" and "position" of the image may be regarded as the listener's uncertainty of the source position—like a probability distribution—and so we measure this uncertainty. The two phones, right and left, are connected to two heads on a special tape-recorder machine, between which accurately known time-delays are introduced. Such time-delays shift the sound image to the right or left of the listener's field and for each setting of delay he is required to guess on which side he hears the image: *right* or *left*. To overcome his personal bias, or inclination to cheat, the successive time-delays are set to random values.

The resulting judgment curve is shown in Fig. 1, using a source of white noise.[6]

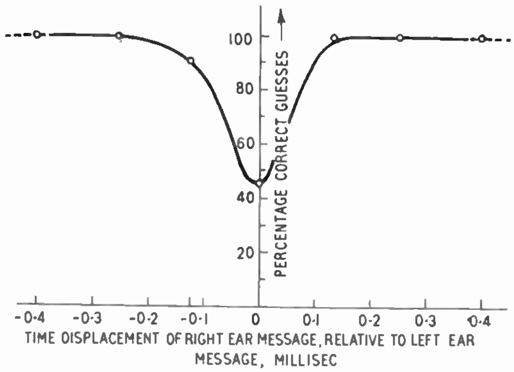


Fig. 1. Right/left judgment curve, using a white noise source of sound.†

Well to each side the listener's guesses are 100% correct; in the middle region he is uncertain for two reasons: (1) the finite subjective "size" of the image and (2) uncertainty as to where his centre is (his nose). The dip here measures his uncertainty, as a standard deviation. This Gaussian-type dip is typical of a random source of sound; an almost exactly similar judgment curve results if we use a speech source instead.

In contrast, if we use a periodic source, like a 1000 c/s tone, the judgment curve itself is periodic[6] as seen is Fig. 2. The "width" of this is indefinitely great, which is consistent with our real life experience, that whereas wideband sources of sound (like noise or sharp clicks) have very definite subjective spacial location, pure tones apparently have none.

These two judgment curves represent extremes, typical of random sources and periodic sources respectively. Further measurement upon speech sources[7] shows that the brain observes that the random breath noises are similar in each ear, that is, highly correlated and it is these which provide the parameters to operate the fusion mechanism. We can now illustrate rather beautifully the process of adaptation to changing circumstances, which I discussed earlier; thus we can destroy the correlation

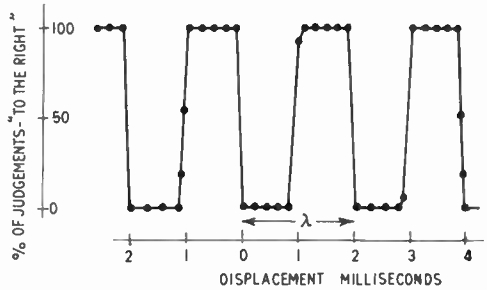


Fig. 2. Right/left judgment curve, using a 1000 c/s pure tone—plotted here as % judgments "to the right".

between the random breath noises in each ear by injecting an independent source of random noise to one ear only. Now the judgment curve is different,[6] and shows sharp dips, as for a periodic source (Fig. 3). The periodic time of the dips here (for the steady sung vowel *AH!*) correspond to the vocal resonances or formants of this vowel. The brain, then, finding its search for common breath noises frustrated, seizes on alternative parameters—the common resonance tones.[7]

We can similarly take judgment curves when two sound sources are present simultaneously—say two people speaking at the same time. If one voice be placed centrally (zero time delay between the ears) whilst the other is moved randomly about it from side to side, the listener's judgments about the position of this latter source are exactly as they were in our first slide—that is, unaffected by the presence of the fixed interfering voice. In other words the images are totally separated by the brain.[8]

To illustrate the essential use of our speech habits, as prior probabilities, to this inferential process of separation, we use two sources now of which the listener has no prior knowledge whatever—namely two independent random noise sources. Measuring the judgment curve exactly as before we find its standard deviation is now *doubled*. Figure 4 shows this doubling. The two sources are totally inseparable and the listener can judge movement only by the shift of

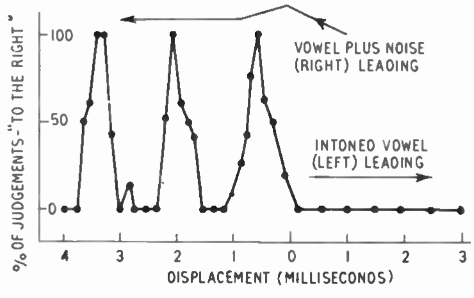


Fig. 3. Right/left judgment curve, using an intoned vowel sound (*AH!*), but with some white noise added to the right ear.

† This diagram, together with Figs. 2, 3, 4, 5, 6, and 7 are reproduced here with kind permission of the Editor of *The Journal of the Acoustical Society of America.*

the centre of gravity of the mixture. They have no separate identities.[11]

I should like to conclude this address with some notes about our binaural hearing in real life—how we use our two ears as we walk about this world and listen. As I said earlier, little advantage of our faculties have yet been taken by any telecommunication system.

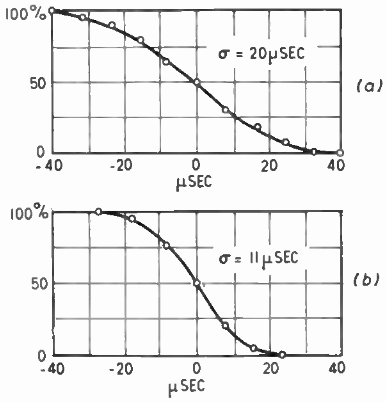


Fig. 4. Assessment of the relative (binaural) separability of (*a*) two independent white-noise sources and (*b*) two voices (same speaker). (The ordinates represent percentage judgments "to the right"; the abscissae represent interaural time-delay, positive indicating left signal leading.)

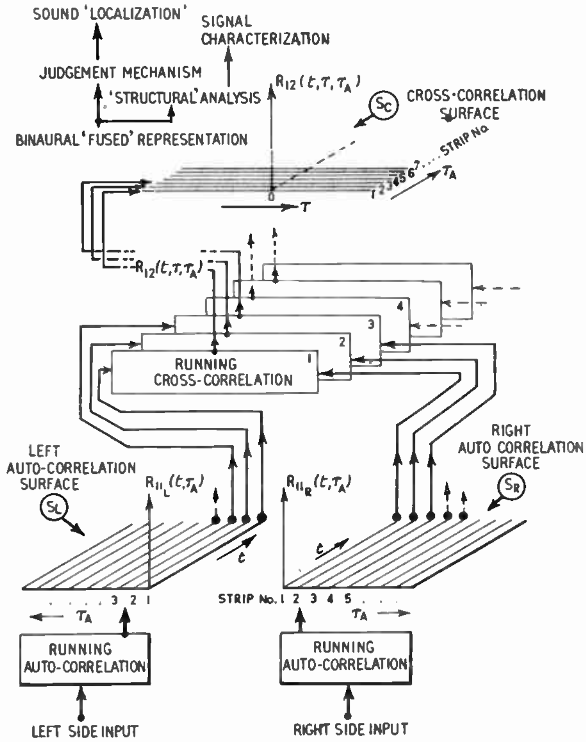


Fig. 5. Schematic of the mathematical model used for computing right/left judgment curves, of which Figs. 6 and 7 show typical examples.

Our principal faculty, I have stressed already, is an enhanced ability to separate different speakers, instruments, or other sources of sound. This results in subjective projection of these sources into different "directions", a process we do not yet fully understand. But essential to this understanding comes, first, understanding of the binaural fusion process itself.

After many hundreds of experiments of the kind I have already described, measuring right/left judgment curves, we have managed to build a mathematical model of this binaural fusion process.[7] I cannot go into all the details here but Fig. 5 shows a kind of "block schematic" version of this model.[7]

The process is one of cross-correlation between the signals reaching the right and left ears—but, in addition, the signals at each ear are each, separately, auto-correlated; one advantage of this auto-correlation is that all-important result—the better separation of images, *gestalten*.

Now the only purpose of making a mathematical model, that is, a theory, is to calculate something with it.[7] In Fig. 6 the full lines show measured results—a right/left judgment curve, using as the source of sound two simultaneous sine waves of frequency 600 and 800 c/s and the dotted curve has been calculated from a knowledge of the signal spectra. Again, Fig. 7 shows the results for an intoned vowel (*AH!*) with noise added to the signal at one ear only.[7]

Such a mathematical model fits all the experimental data we have observed, if the listener wears headphones.[7] But for real life, external, projected sound impressions, we listen unfettered by headphones. To assess sound directions we turn our heads about, or tilt them this way and that, getting many varied samples for the brain to get to work on.

The sounds reaching our two ears differ in several ways, not only in time, as many people imagine, but in relative mean intensity too and, quite important, in short-term spectra quality due to the shape of our heads and ears. And echoes also play their part.

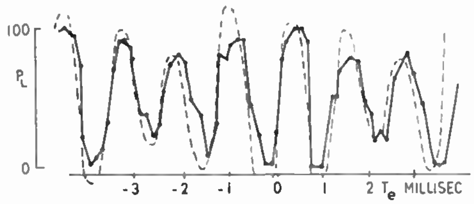


Fig. 6. Measured and computed right/left judgment curves, using as a source two sine waves of 800 c/s and 600 c/s.

This brings me to the final part of my address, a part which I personally feel to be of the most significance and, if I do nothing else at this Inaugural Meeting, I want to convince you now that we do not understand fully how our stereophonic sound systems (using two loudspeakers) work. As I go about among sound engineers, or listen to lectures, or symposia on stereophony, I am struck not only by the inadequacy of the theory used, as a basis of design, but often by the fact that it is completely wrong.

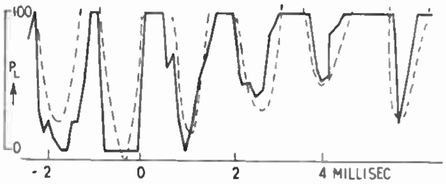


Fig. 7. Measured and computed right/left judgment curves, using as a source the intoned vowel *AH!* noise-masked at one ear only (S/N ratio—15 dB).

First, let us look at the geometry of the business (Fig. 8). Here we have two microphones in the studio (either spaced, or crossed, depending on the system) connected, effectively, to two loudspeakers in your sitting room. Now, in contrast to wearing headphones, with each earphone feeding one ear only, *each* of the loudspeakers here feeds *both* your ears—that is, *four* sound rays. It is often stated that the two microphones in the studio simulate the two ears of the listener; but how can this possibly be the case? It is of course nonsense! For one thing, if you turn your head, the microphones don't move.

The only really adequate theory of this situation of which I know is in a London University Ph.D. thesis by Dr. D. M. Leakey,[9] part of which has been published.[10] He includes all ways of head turning and tilting.

As the listener moves his head about, so the signals at his ears vary in several ways. If the speaker in the studio walks about, so the listener may have the impression of movement. But there is no reason whatever why this illusion should correspond with the truth!

Further than this, if there are several sources in the studio, we cannot use such simple geometry for each one—unless we can first understand how they are kept as separate *gestalten*—which we cannot, yet. Again, although the listener may hear the instruments spaced about, they may have gone into different positions by the time his illusions are set up. For the subjective position of any instrument is partly a function of its sound quality.

I do not have any intention of being disparaging, in this subject, a favourite one of mine. For good, usable, satisfying, systems of stereophony have been built. Then engineers often get there first, before the theoretician. But we don't *really* understand why they work —nor do we yet understand many of the faculties of human listeners, some of which I have referred to tonight. But I think it very important that we should try to understand better, by doing more experiments on people, and less on hardware. On such research a great deal of the future successes hang—such as will be described and discussed at the future meetings of this Electro-Acoustics Group.

And what will be these systems, which will be described and discussed? I cannot say, of course; otherwise there would be no point in my referring to lines of research, as I have done tonight. But some of you may be a little impatient of talk about long-term research and would want more concrete definite suggestions as to what results might accrue from work of the type I have indicated. In other words, can I suggest practical proposals?

I can, and will quickly indicate about ten; none of these are yet completely clear and none are definite inventions.

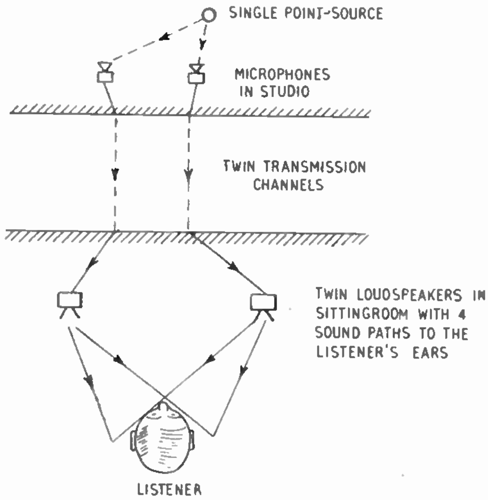


Fig. 8. The simple geometry of two-loudspeaker stereophony systems with spaced microphones, when listening to a single point source.

The first, and most obvious, is the application of better understanding of perception processes to compressed speech channels. We already have the Vocoder and the Lawrence-type[1] of parametric system. Neither are perfect and both can be advanced by better understanding of human hearing, especially of the true parameters of speech which set up aural images, *gestalten*, under different noise conditions. In this connection, it is interesting to note the psychological difference between these two broad approaches to speech compression. Thus the Vocoder imposes a fixed set of filters upon the speech signals and therefore introduces what we call *distortion*—small perhaps but nevertheless distortion. On the other hand the parametric systems introduce varying errors, but errors which, instant by instant, depend upon the structure of the speech signals themselves; the distortion here has more the nature of an *accent*. Psychologically, accent and distortion are quite different. We need to know far more yet about the mental processes involved when dealing with both departures from our own speech norms.

Next, most of our communication channels have fixed characteristics (bandwidth, and so on) through which we send our speech signals, together with whatever noise happens to exist. If we understood better the psycho-physical processes of voice and noise formation we might control these transmission characteristics so as to maximize the listener's chance of forming the voice image and excluding the noise image. This is an optimization, not in the Wiener filtering sense of maximizing the r.m.s. signal/noise ratio, but of the *perceptual* chance of success.

Closely related to this is a possible application of binaural listening to improve the subjective speech signal in relation to the noise. If we knew more of the manner in which the brain uses the sounds reaching our two ears so as to increase our discrimination between sound images we might envisage a double-modulation channel, perhaps with headphone listening, in which the signal-to-noise relation is improved, not in the conventional sense of a simple decibel ratio but so as to cause the speech signal to go over *there* and the *noise* image to form over *there*.

Another, but well-known, application of better psychological understanding would be to the so-called "speech typewriter." Basically, the idea here is to set up a machine which recognizes the phonemes of speech; these the machine could type down in a special script or it could transmit them over communication channels of very reduced capacity. As most people know the great problem here is that no two people speak exactly alike. Nevertheless extraordinary success has been obtained, both in the U.S.A. and here especially, in my view, with the machine of Professor Denis Fry at University College, London.[12] His machine has a built-in store of probabilities of speech sounds, which helps the machine to make the right guesses.

Strictly speaking, this problem is an insoluble one, but we might compare it to the problem of machine translation, which is also impossible. In both cases a practical compromise is sought, because this still can be of great value. Both have a human being at the output end who can supply the necessary editing processes. But we can go a long way yet in making his task easier.

On another plane altogether, our psychological studies may eventually lead to far more meaningful specifications and criteria, by which we assess the qualities of our channels. For example, most engineers speak of "signal/noise ratios", or of "cross-talk ratios", as though a single figure, like 40 dB, can tell us how good the channel will be. But the mental separation of signal from noise, or of two speech signals depends not on a simple ratio question, but is a psychological question involving the micro-structures of the signals and also the listener's prior knowledge of them. I have tried to give some notion of this problem tonight; perhaps better specifications will eventually emerge from applying some of the methods of perceptual measurement, such as I have indicated.

There are other possible practical gains to be had, from psychophysical experiments on which we might speculate. Understanding of brain-like activity, including perception, is already revealing processes which hitherto we had thought to be beyond any machine. Our whole concept of machines and their capabilities has had to be revised. Processes of learning and adaptation, of recognition and decision making, and other animal-like actions, can now be defined mathematically—and so instrumented. One great difference between a brain and a classical machine lies in the size of the memory store. Machines in the past could act only on immediate instructions. Modern machines with stores can act according to their past experience also. Not only have brains memory-stores of astronomical scale but they possess other faculties not yet understood and so not imitated by machines. For example, in spite of the vast scale of the human brain-store, we seem to have *access* to any part of it in fantastically short time. We cannot yet conceive how such access can be structured.

I could go on and give many examples. These examples are not, of course, inventions yet! They merely give my own views about lines of thought, and of psychological experiment, which it might be fruitful to pursue and which will lead to invention.

I, myself, have no doubt that, in the future of this Group of your Institution, the papers which will be read and discussed will contain an increasing reference to human beings. We shall be forced more and more against the limits of our physical knowledge and the attention of more and more engineers will be brought to the real nature of the communication process—that communication always, at some stage, involves people.

References

1. W. Lawrence, "The Synthesis of Speech from Signals which have a Low Information Rate", "Communication Theory", p. 460 (Butterworths, London, 1953). (Paper read at Symposium on "Applications of Communication Theory" London, 1952.)

2. E. Colin Cherry, "Some experiments upon the recognition of speech, with one and with two ears", *J. Acoust. Soc. Amer.*, **25**, No. 5, pp. 975–979, September 1953.
3. E. Colin Cherry, B. McA. Sayers, and P. M. Marland, "Experiments upon the total inhibition of stammering by external control, and some clinical results", *J. Psychosomatic Research*, **1**, pp. 233–246, 1956.
4. G. A. Miller, "Language and Communication" (McGraw-Hill, New York, 1951).
5. D. Middleton, "An Introduction to Statistical Communication Theory", (McGraw-Hill, New York, 1960).
6. E. Colin Cherry and B. McA. Sayers, "The 'human cross-correlator'—a technique for measuring certain parameters of speech perception", *J. Acoust. Soc. Amer.*, **28**, No. 5, pp. 889–895, September, 1956.
7. E. Colin Cherry and B. McA. Sayers, "Mechanism of binaural fusion in the hearing of speech", *J. Acoust. Soc. Amer.*, **29**, No. 9, pp. 973–987, September, 1957.
8. J. A. Bowles, "Binaural Discriminatory Processes", University of London M.Sc. Thesis, 1960.
9. D. M. Leakey, "Investigations into Stereophonic Hearing in Communication Channels", 1960. Faculty of Engineering University of London Ph.D. Thesis.
10. D. M. Leakey, "Some measurements on the effects of inter-channel intensity and time differences in two channel sound systems", *J. Acoust. Soc. Amer.*, **31**, pp. 977–86, July 1959.
11. E. Colin Cherry and J. A. Bowles, "Contribution to a study of the 'cocktail party problem'", *J. Acoust. Soc. Amer.*, **32**, p. 884, July 1960.
12. D. B. Fry, "Theoretical aspects of mechanical speech recognition", and P. Denes, "The design and operation of the mechanical speech recognizer at University College London", *J. Brit. I.R.E.*, **19**, pp. 211–34, April 1959.

Manuscript received 20th October 1960 (Address No. 24).

DISCUSSION

Under the chairmanship of Mr. H. V. Leak (Member)

Mr. K. R. McLachlan (*Associate Member*): May I, first of all, thank Professor Cherry on behalf of the Electro-Acoustics Group for a most lucid and interesting exposition of the work on hearing which he and his colleagues have been pursuing.

There is one point on which some clarification may be needed, and it concerns Professor Cherry's early comment that subjective rather than objective experiments are of greater ultimate value. Whilst all experiments may be regarded as subjective since they involve human operators, to many engineers in the field of audio frequency engineering the term subjective testing means using the senses directly for assessing the performance of a system. Conversely, objective testing implies the interposition of measuring equipment to reduce to a minimum the dependence on sensory information and in particular to reduce this dependence to one of the direct comparison of simultaneously presented data. It would be most helpful therefore, if Professor Cherry would elaborate on his definition of the words objective and subjective.

Whilst appreciating that the work described has been directed towards obtaining a satisfactory model on which the various aspects of hearing in the human being may be explained, it would be of interest to know if he has conducted any experiments with, say, combined aural and visual stimulation. This would be particularly relevant in the "cocktail party" effect and other similar situations in which the selection of a desired signal in the presence of unwanted signals takes place.

In the "shadowing" experiments Professor Cherry has described how the subjects are influenced by their normal speech habits, to the extent that an American "object" will not produce an American "shadow" in an English subject. This, superficially, would seem to contradict the statement that a subject who normally stammers can "shadow" a "non-stammering" voice without difficulty.

Professor Cherry (*in reply*): I think Mr. McLachlan has slightly misunderstood my comparison of subjective and objective results. I agree completely with his own view, in fact. To me "objective" means "predictable and repeatable". Even if a "measuring instrument" is employed, the reading of this is a subjective experience too; but it provides "reliable, repeatable" *comparison* of two subjective sensations (e.g. the sight of a needle and of a scale gradation). All I want to stress is that good *objective* measurements of subjective phenomena can be made, using good techniques of experimental psychology, but that many engineers seem not to realize this and feel apologetic about subjective work.

I have done no experiments with simultaneous aural and visual stimulation; certainly, we use both at cocktail parties!

A subject who stammers can "shadow" a non-stammering control, without stammering, provided that the two people share a common language and dialect (i.e. common syntactic habits). American and English syntaxes are markedly different, and the point of interest is that the "shadower" adheres to his own habits of syntax.

Mr. L. Nelson-Jones (*Associate Member*): Why is it that, in my experience, when listening to information in the presence of other interfering noises, one often understands much more when listening casually than when listening with deliberate intent to pick out the required information; or is this an illusion?

Professor Colin Cherry (*in reply*): It depends upon the intellectual level of the material you are attending to. In casual conversation you may respond very readily by habit (which is deep in all of us) and the "higher" centres of the brain may be little involved. But introspection is a very deceptive thing; what you think you do and what you really do may be very different. Your degree of concentration can be controlled in behavioural experiments, and what you really do can be measured.

Mr. J. A. Cole: Water waste inspectors use primitive "stethoscopes" to detect leak sounds at night-time. In this task they not only require good hearing but an ability to distinguish a leak noise from that due to distant traffic. Thus a man possessed of good intelligence and a good memory for sounds will succeed at this task better than another with hearing just as good, or even better than the first man's, but with poorer mental abilities. This observation may add to Professor Cherry's remarks on the need to know more about the abilities of human listeners to discriminate sounds.

The "shadowing" experiment described by Professor Cherry has a close parallel in the calling of cock Red Grouse, who reply to one another's alarm notes in unison.

Professor Cherry (*in reply*): "Perception" is of course not solely a question of the physical sensitivity of the reception organs, the ears, eyes, etc. It depends upon the perceiver's expectations as built up from his whole lifetime's experience. Certainly, training will greatly improve the faculty of separating one source of sound from another.

The calling, and imitative response, of the cock Red Grouse is an interesting example of *mimicry*, which is a function operating widely in aural perception, both human and animal. It plays a vital part in child learning of speech and, in my opinion, it continues as a learning function throughout our lives. I regard mimicry as one of our fundamental faculties.

Mr. Gordon Hyde: In the field of vision it is now well established that there is a perceptual space, acquired by the individual through his contact with a rectilinear universe. It is also well established that this perceptual space is not a property of the nervous system at birth, but is acquired by a learning process. Is it not possible that acoustic perception is also a learned phenomena, related to the organism's exposure to acoustic stimuli during the early stages of development? This appears to raise interesting questions regarding auditory perception in later life. To what extent can one claim that even precise measurable parameters are not influenced by such phenomena. To what extent also do specialized listening habits, such as for instance the widely different musical traditions of Europe and Asia, influence the fundamental parameters of hearing?

Professor Cherry (*in reply*): To me there are not two worlds—"perceptual space" and "real space"—but there is a world for every person. Inasmuch as these worlds overlap we can communicate and share our knowledge. My world, apart from the limited reflexes built in me at birth, has been acquired as a "model", developed from my lifetime of sensory experience. It changes every day.

Thus what you will see, hear, smell, etc., depend, almost entirely, upon your past; depend too upon your cultural background and traditions.

Mr. Hyde (*Communicated*): It had not been my intention to imply that the more complex phenomena of acoustic perception would be susceptible to investigation with the objective rigour of Professor Cherry's own work. It was my intention to imply that the practical man in the electro-acoustic field is asked to deal with these complex phenomena. Professor Cherry himself has drawn attention to the fact that current theories of stereophony do not even have a subjective correlation to the realities of acoustic perception. Clearly, in the present state of the art, even quasi-empirical knowledge would be better than nothing at all and may even suggest fruitful lines for more rigorous investigation. It will be hoped that the new group will bring to our attention both the results of rigorous research, such as that described by Professor Cherry, and the advances made by engineers and creative people applying electronic knowledge to acoustic problems in a more empirical way.

Professor Cherry (*in reply*): I think it vitally important that engineers, engaged in designing systems like television and stereophony, to tickle our sensory surfaces, should have some understanding of psychological processes, because they are (usually) raised in the tradition of physics. It is not that physics is wrong, of course, but that its language and concepts are not enough. It deals only with the overlapping parts of our private "models" we call the "real world". But engineers are for ever forced to make compromises, for economic reasons, and in so doing run the risk of producing bad illusions, contravening our "models". Briefly, they make their systems for *people* and they must know what makes people tick.

Mr. C. T. Chapman (*Member*): Although the fundamental frequency of the diaphone fitted to lightships for aid to vessels in fog is very low indeed, it starts very abruptly and finishes abruptly after a very steep rise in frequency. It is probably this distortion that enables quite an accurate bearing to be taken using only the ears for location of direction.

Professor Cherry (*in reply*): I believe the very low frequency is used to aid long-distance propagation of sound. But I am sure it is the abrupt start and stop which give the clues to direction.

Mr. R. L. West (*Member*): I too am inclined to agree with the last speaker and quote a wartime episode that stands out in my memory. I was walking in the quiet countryside amongst the Chiltern Hills and distinctly heard faint and distant "whoomphs". We were not to know for several days that this was the Dunkirk evacuation fighting. The direction was quite unmistakeable every time I was in unobstructed surroundings. The sound level was not very far above threshold as it needed temporary cessation of breathing to be sure it was real and not imagined.

Back in town on the following days the bombardment could still be heard, even indoors, with enough concentration, but there was of course then no directional effect. The use of the word "whoomph" was to suggest a non-sinusoidal sound containing very low frequency components only, and not having a very abrupt start or finish. The distance must have been some 140 miles. I also remember quite distinctly that the sound seemed to come from some 15 deg—20 deg above the horizon. The impression at the time was of something very loud but not very far away.

Professor Cherry (*in reply*): I wonder how abrupt were the start and finish of these distant explosions? Don't forget that the main *energy* might be centred around the low frequency region, but that the sounds might be started and stopped in relatively short time. Again, in sensing *direction*, the brain will make use of every available transient sound. I have not seen the waveforms of distant explosions, but I would guess them to be, not sinusoidal (what is resonating?) but irregular, with very many small transient components.

Mr. J. A. Le Warne (*Graduate*): I have listened with interest to the results of experiments on the "apparent location" of sounds by time spacing audio signals into separate headphones. Can you please give any results of experiments carried out on reception of sound by bone-conduction?

Secondly, if such experiments have been carried out, what were the results of positioning transducers at various places on the skull?

Thirdly, could use be made of bone conducted sound in "locating" an information source in conditions of high ambient noise?

Professor Cherry (*in reply*): Bone-conducted sound is of vital importance to the self-perception, and monitoring, of your own voice when speaking. Again we have found that, if we repeat the well-known experiment of producing artificial "stuttering" and drawl by delayed playback of one's own speech into the ears, but with the headphones pressed on to the *mastoid* region (with headphone covers unscrewed) behind the ears, the subjective sensation is vastly more distressing.

We have done no work on sound source location using bone conduction, and I know of no way of improving our directional sense in this way in high noise conditions.

Mr. J. N. Holmes: Could Professor Cherry say whether a person performing a "shadowing" task takes in any of the meaning of what he is saying?

Professor Cherry (*in reply*): Very little, if anything, of "long-term" content. Short-range meanings are possible to extract, but a long argument or description could not be followed.

If we "shadow" one message source applied to one ear, whilst a totally different one is applied to the other ear, then nothing whatever is perceived of this non-shadowed source—not even the language. But it is usually known whether it is a human voice or not.

Mr. A. I. Forbes Simpson (*Member*): I should like to quote from experience of the explosive type of phenomenon associated with "bump testing": that one perceives this phenomenon with one's "stomach" (more accurately I expect at the diaphragm) and so, although no discrimination could presumably occur aurally, one has a second source of information which may explain matters.

I should like to ask Professor Cherry if the automatic responses discovered during the "shadowing" tests do in fact occur with normal listening, i.e. is one in fact producing the muscular responses of speech when listening? In this connection I would mention the case of the opera singer with throat trouble who did not recover until she was debarred from playing the piano.

Professor Cherry (*in reply*): I agree that we should not dismiss our kinesthetic sense when considering our directional faculties. It may well be that we detect directions of very low frequency sounds this way, possibly even if sinusoidal.

Regarding "shadowing" in normal life, we normally inhibit actual muscular action, but I am thinking that the neural centres might be involved, sub-threshold. Some people *do* mutter whilst they listen. The opera singer story is a good one; there are many sources of evidence that our hearing and speech are closely integrated. There is the common experience of holding one's breath when listening to a long sustained note by a singer. I, personally, believe speaking and hearing to be so integrated as to be inseparable; we should study them as one whole process.

Mr. J. E. Burns: The presence of pairs of correlating networks in the mathematical model makes one wonder whether there is any obvious physiological process (other than in the brain) to which they would correspond. Some modern theories on pitch discrimination assume that there are two processes, one for low frequencies and the other for high. For low frequencies the ear is supposed to change the vibrations in the air into the same frequencies of impulses along the nerves originating in the cochlea. Thus if a low-frequency tone is incident upon both ears, the nerve impulses reaching the brain from each ear will be highly correlated. High frequencies on the other hand are supposed to be discriminated by their position of resonance along the basilar membrane, thus exciting only those nerves originating near that point. The frequency of nerve impulses will in this case bear no relation to the frequency of the sound, and the impulses reaching the brain from each ear will be completely uncorrelated. This may be an explanation of why high frequency sounds are easy to localize by binaural localization, whereas low frequency sounds are not.

Professor Cherry (*in reply*): The mathematical model of the *fusion* process (not of pitch discrimination) we put forward is a model of behaviour, not of physiology. Thus it is "one of a class of possible models", not a unique model; an infinity of transformations are conceivably possible.

Nevertheless, in this fusion process too, low and high frequency sounds are treated differently. Whereas low frequency sounds are fused into a single image, only the *envelopes* of high frequency sounds are fused. If the envelopes are non-existent (constant, steady, amplitude) no fusion occurs.

A simple experiment illustrates this. If a sound of, say, 2000 c/s is played to one ear through a headphone, and 2100 c/s to the other ear, no fused image forms. If then they are given a *common* modulating envelope of, say, 400 c/s they immediately fuse. Undoubtedly this low/high frequency discrimination in fusion plays an important part in directional hearing. I would add that it is possible (but not certain) that our *auto*-correlation process may occur in the ear itself, perhaps in the cochlea, and it is this process which extracts the envelope of sounds.

Mr. H. F. Lloyd: One would expect to get maximum correlation between the signals received at the two ears, from a particular source, when the listener is directly facing that source. In a typical "cocktail party" situation, however, one's experience is that to concentrate on one speaker out of many one does not in fact hold the two ears symmetrically to that speaker. Can this be explained by the mathematical model?

Professor Cherry (*in reply*): We have often performed the experiment of sitting a listener inside a small tent, requiring him to respond to one speaker when another is placed diametrically opposite. Watched in a mirror overhead he is always seen to sit with his ears in line with the two speakers, never facing the wanted one.

The full mathematical model is not required to explain this. Let $f_1(t)$ and $f_2(t)$ be the two different speakers, and let T be the time delay of sound travelling the distance between the two ears. Then, sitting in this way:

Total sound reaching left ear $= f_1(t) + f_2(t - T)$
Total sound reaching right ear $= f_1(t - T) + f_2(t)$

Suppose the brain tests various hypotheses about T (assuming its value is not built-in the brain) by inserting relative time delays $T_1, T_2, \ldots T \ldots$ in turn. When it strikes the correct value T the signals, above, are now:

Left: $f_1(t + \frac{1}{2}T) + f_2(t - \frac{1}{2}T)$ ($\frac{1}{2}T$ added)
Right: $f_1(t - \frac{3}{2}T) + f_2(t - \frac{1}{2}T)$ ($\frac{1}{2}T$ subtracted).

Simple subtraction now eliminates one source, f_2, leaving f_1 and a slightly delayed "echo" also.

If the speaker *faced* one source, $T = 0$ and no such separation is possible.

It is such "hypothesis" testing which, if made more realistic than this simple description, leads to our mathematical correlation model.

Mr. W. M. Dalton (*Associate Member*): Professor Cherry makes it sound too easy. A sound will give the same time difference at the two ears when coming from any position around a circle. How then does one locate up and down?

May I add further that an aeroplane to the right could be at 45 deg or 135 deg from the direction of the noise. How does one know which? The answer might be that we do not know until we turn to look; then in turning the ear passes through 90 deg which gives a maximum of difference.

Professor Cherry (*in reply*): On the contrary, I am on Mr. Dalton's side! I certainly don't want to imply that sound *direction* is assessable from interaural time-interval only. Other cues are used; for example, interaural amplitude differences, small spectral differences, and others, all overshadowed by the effect of our expectations (past experience of positions of similar sources) and sense data from the eyes and perhaps other organs.

Our sense of direction in the vertical plane is poorer than in the horizontal; aeroplanes are most difficult to locate, unaided. The discrimination between 45 deg and 135 deg can be executed by turning or tilting of the head. Only very small movements are required, in theory, but aeroplane sounds are rather "steady-state" and therefore difficult to deal with.

Mr. M. B. Martin (*Associate Member*): In view of the described behaviour of subjects "shadowing" speech could Professor Cherry say how people carrying out simultaneous interpretation perform their task?

Professor Cherry (*in reply*): Simultaneous-translators provide a very complex psychological study. They have of course two sets of engrained motor habits and I am sure, nevertheless, that they are (sub-threshold) "shadowing" the voice of the speaker. Their actual translated version is never truly simultaneous, and cannot be, for the two languages are time-structured in different ways. I think that the situation is akin to that of listening to two conversations "at once", at a cocktail party. We don't do it "at once", of course, but sample and guess wildly by inference, using the enormous redundancy in conversational speech.

I once heard a simultaneous translator, struggling to interpret a German speaker, give up at one stage for nearly 20 seconds, then gasp out into his microphone: "the verb man, for God's sake, the verb!"

News from the Sections in Great Britain

Members in Great Britain have already received the programme of Institution meetings to be held in the ten Sections during the second half of the current session (January to May 1961). The programme lists details of over fifty meetings and there are in addition six symposiums to be held in London, each of which will include the presentation of several papers.

Section Committees are being most helpful in providing the Editor with information regarding their local activities. It is hoped to give more space in future issues of the Journal *to these reports.*

Southern Section

The first members meeting of the newly-formed Southern Section took place on Wednesday, 26th October 1960, in the Lanchester Building, Southampton University. In his Chairman's Address, Commander J. S. Brooks, R.N. (Associate Member) developed as his main theme variety reduction in components, their usage and storage. Naturally, Commander Brooks was especially concerned with the logistic and storage problems peculiar to H.M. ships and in particular the rationalization of common electronic component spares carried on board. A very lively discussion ensued which ranged over the field of electronic component standardization and reliability.

Local members are very interested in all forms of navigational aids for both sea and air. The second meeting was a representation of a paper given before the Radar Group in London. This meeting was held on 22nd November in Farnborough Technical College, when Mr. J. S. Shayler, B.Sc.(Tech.) presented the paper published on pages 17 to 33 of this *Journal.*

On Wednesday, 14th December 1960, at Southampton University, Mr. A. E. Crawford read a paper outlining "Some New Piezoelectric Devices".† He first explained the differences between single crystal materials and the polycrystallic ceramics, then described how the latter material was employed in various devices and went on to discuss possible applications. Included in this survey were transformers, filters, transducers, strain gauges, ignition systems and generators. The paper was well illustrated with slides and was also enlivened by some demonstrations, one of the most effective being a simple gas lighter. The large attendance showed the considerable interest in piezoelectric devices.

F. E. P.

South Midlands Section

About 45 members and visitors attended the twenty-ninth meeting held by the Section since it was established in 1956. The meeting was held in the Malvern Winter Gardens on 3rd November, when Professor D. G. Tucker (Member) and Dr. D. E. N. Davies presented a joint paper on "Electronic Sector Scanning in Sonar and Radar Systems".

Professor Tucker described the operation of a sonar sector scanning system, and pointed out that sector scanning was more important in sonar than in radar systems due to the low velocity of sound waves in water. In principle, an array of receiving elements was arranged with swept-frequency frequency chargers coupling each element to a delay line, so that the receiver effectively "looked" over the whole sector which was being scanned within the duration of the transmitted pulse. Details of a multiplicative array giving half the beam width of an additive array were described.

Dr. Davies then described his work in applying these principles to radar sector scanning. The special problems associated with radar systems were discussed, including the distortion produced by increasing the scanning rate.

The discussion ranged over a wide variety of topics, including questions of i.f. bandwidth in the radar system, signal/noise ratio due to the large numbers of mixers employed, effects of aeration in sonar systems, and Doppler effects.

Editorial Note: Dr. Davies will read a paper on "A Fast Electronically Scanned Radar Receiving System" at a meeting of the Radar and Navigational Aids Group in London on 1st February 1961.

A. H. M.

West Midlands Section

Members who attended the meeting on 26th October, held in Birmingham University, heard a most interesting address by Mr. R. J. F. Howard on the subject of automatic control of industrial processes.

The author gave many examples of automatic control systems, referring first to those for the accurate control of motor speed. These included methods for preserving a constant relationship between the speeds of several motors—important in the synchronization of conveyor belts in a large plant.

Examples of control systems for operating a complete process were given, ranging from a simple timing device to apparatus used to control the centrifuging of sugar. Position control systems were then discussed, among them the automatic control of machine tools to reproduce a given profile, the remote control of irradiation equipment in radio-therapy, and the use of photocells to ensure correct registration of the three colours used in colour printing. D. H. A.

† This paper, first presented at the Symposium on New Components in London in October, will be published in an early issue of the *Journal.*

U.D.C. 621.396.933.23

Radio Guidance Elements of the B.L.E.U. Automatic Landing System for Aircraft

Presented at a meeting of the Radar and Navigational Aids Group, held in London on 11th May, 1960.

By

J. S. SHAYLER,

B.Sc. TECH., (*Associate Member*)†

Summary: The object of this paper is to describe the radio aids which form part of the B.L.E.U. automatic landing system and only a brief description is given of the system itself. The internationally accepted i.l.s. is used as guidance during the approach phase, and reasons are given why this cannot be used to complete the landing. Two new aids are therefore introduced for the landing phase, magnetic leader cable for azimuth guidance and an f.m. radio altimeter for height guidance. These equipments are described in turn in some detail. Finally there is a discussion on the two main problems to be solved in introducing the system to civil aviation, namely ensuring the extremely high safety level required and finding a guidance system whose ground elements will, at the same time as satisfying the accuracy requirements, be completely acceptable to the installing authorities.

1. Introduction

The automatic landing system developed by the Blind Landing Experimental Unit (B.L.E.U.) of the Royal Aircraft Establishment has already been broadly described on a number of occasions.[1, 2, 3] There is, however, no detailed description of the guidance components, and it is the purpose of this paper to rectify this, as well as to discuss some of the major guidance problems to be solved in introducing the system to civil aviation.

2. The Reasons for Automatic Landing

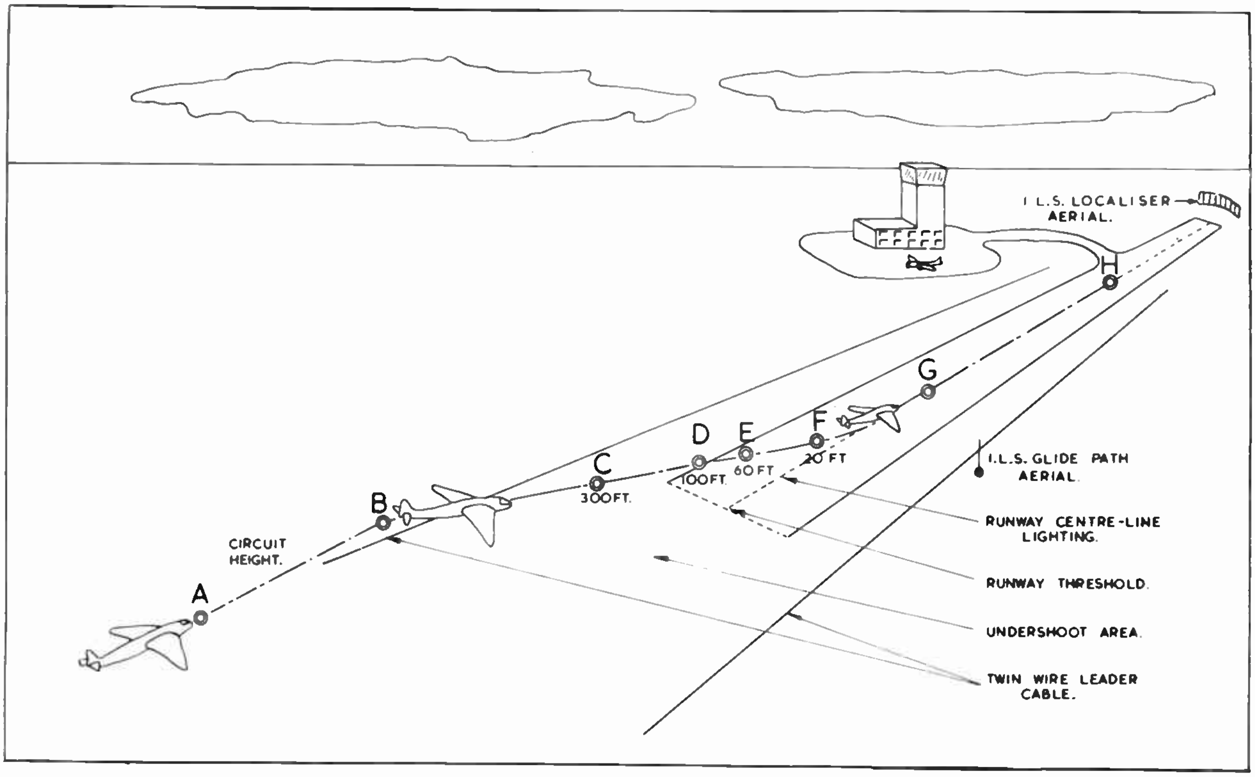
Since aircraft first flew there has been a steady improvement in instruments and radio aids so that today bad-weather flying is a commonplace and safe procedure. In fact the most restrictive phase in bad visibility is now that of landing, and by landing is meant the descent from a height of about 200 ft to touch-down of the wheels on the runway.

There are in use a number of aids which allow a pilot to approach the runway under Instrument Flying Rules down to a height of about 200 ft, but here, in order to complete a landing, he must be able to interpret visual clues from the ground and make a normal visual landing. If he cannot see these visual clues, the pilot must overshoot and divert to another airfield where visibility is better.

The minimum height allowed on instruments is determined by the time the pilot takes to perform the landing manoeuvre, which includes interpreting his visual clues, correcting any lateral error from the runway centre-line that remains at the end of his instrument approach, and finally assessing any height error and performing the flare (i.e. reducing the rate of descent) prior to touch-down. Many tests have been made using very different types of aircraft[4] and the minimum time for the whole sequence is about 15 seconds, with 20 seconds being more realistic. Since a typical rate of descent is 10 ft per second, the minimum height of about 200 ft for starting a visual landing is obtained. Although with some aircraft in some conditions this height may be reduced a little, it seems unrealistic to expect a major reduction where the final touch-down must be made visually. The requirement for a fully blind landing system therefore becomes apparent.

There are two main categories of blind landing systems, one in which the pilot is given sufficient information on his instrument display for him to land the aircraft manually, and the other in which the aircraft is landed automatically, leaving the pilot in the important role of monitor. Work is in progress in B.L.E.U. on both systems, but from experience on automatic approach, since confirmed by work on automatic landing, the automatic system has the two advantages: (i) of removing the strain from the pilot during one of the most difficult phases of flying an aircraft, particularly in bad visibility, and (ii) of giving more consistent touch-downs in any weather conditions than a pilot can maintain even in good visibility. There is therefore a strong argument that an automatic system should be the primary blind landing aid, although there is an unresolved problem, which will not be discussed in this paper, of whether there is a place for a pictorial manual blind landing system as an independent monitor.

† Blind Landing Experimental Unit, Royal Aircraft Establishment, Bedford.



	PHASE	GUIDANCE AZIMUTH	GUIDANCE ELEVATION
A – B	Automatic Approach (Track)	I.L.S. Localizer, Compass	Barometric Height
B – C	Automatic Approach (Glide)	I.L.S. Localizer	I.L.S. Glide Path
C – D	Automatic Land (Leader Cable)	Leader Cable	I.L.S. Glide Path
D – E	Automatic Land (Attitude)	Leader Cable	Fixed Attitude
E – F	Automatic Land (Flare)	Leader Cable	Radio Altimeter
F – G	Automatic Land (Kicking Off Drift)	Compass	Radio Altimeter
G – H	Manual Ground Run	Leader Cable, Compass	

Fig. 1. The Automatic Landing System.

3. System Considerations

The broad considerations that have influenced the choice of the particular B.L.E.U. automatic system will now be briefly discussed.

One of the aids by which a pilot can approach down to a height of 200 ft is the internationally accepted and widely used radio aid i.l.s., (Instrument Landing System—a misnomer),[5, 6] briefly described in Section 6.1. Furthermore, autopilots have been developed and are in current use which enable an aircraft to be coupled to the i.l.s. beams, thus allowing a completely automatic approach to be made from a range of 10 miles or more. This proven system has been made the basis of the automatic landing system, items only being added to enable the final 200 ft height to be traversed. These additions consist on the one hand of two new short-range precision aids, namely Magnetic Leader Cable for azimuth guidance and a Frequency-modulated C.W. Altimeter for height guidance, and on the other hand of modifications to the autopilot computer (coupling unit) to enable it to compute the landing, in addition to the approach, manoeuvres.

A feature of the system is that the computing of the control commands is performed in the aircraft, where aircraft attitude information is already available. This is strongly held to be preferable to a system where the computing is performed on the ground.

4. The Automatic Landing System

In this section the system will be described generally, with the aid of Fig. 1.

The landing starts with a standard automatic approach (A, B, C) in which signals from the i.l.s. localizer (azimuth beams) and i.l.s. glide path (elevation beams) are fed to the automatic pilot to lock the aircraft to the radio beams. However, for reasons to be given in section 6.1, the i.l.s. beams are not suitable for landing, and so the following sequence is introduced.

At a height of about 300 ft (point C) the aircraft comes within the field of the magnetic leader cable and, on receipt of suitable signal strength, the output from the leader cable receiver in the aircraft is substituted in place of the i.l.s. localizer signal as the input into the azimuth channel of the autopilot. With this more precise azimuth information, any lateral error from the runway centre-line can be reduced before touch-down.

At a height of about 100 ft (point D) i.l.s. glide path signal becomes unusable on many sites, so this signal is removed from the autopilot. It is too early, however, to guarantee a good guidance signal from the radio altimeter because the aircraft is not yet over the level runway, and so, for a few seconds, the aircraft attitude is held constant at a value averaged during the whole of the approach to this stage. At a height of about 60 ft (point E) the flare begins. This is the manoeuvre which progressively reduces the rate of descent from about 10 ft per second on the approach to about 2 ft per second at touch-down so that a smooth landing is made. The radio altimeter, which so far has only been used to initiate the switching at preset heights can, during this phase, be introduced as a guidance element, and its signal is coupled into the autopilot channel in such a way that the rate of descent of the aircraft is made proportional to its height, i.e. the lower the aircraft the more slowly it sinks.

The final sequence begins at a height of about 20 ft (point F), when the kicking-off drift manoeuvre is initiated. During this phase the wings are held level, but rudder is applied to remove any drift due to cross-wind, so that the aircraft touches down not only close to the runway centre-line, but also lined up with the runway (point G).

One other automatic loop must be mentioned. During the whole of the approach, the airspeed is held constant by an automatic throttle control that uses airspeed and pitch attitude as input data. This also automatically closes the throttles during the flare so that when the aircraft touches down the engines are at flight idling speed.

After touch-down the pilot, during the ground run, disconnects the autopilot and brings the aircraft to a stop manually (point H), using as guidance either the runway centre-line lighting pattern or the leader cable information displayed on his flight instruments.

During the whole of the approach and landing, guidance error information and manoeuvre demands fed into the autopilot are displayed on an integrated flight instrument display so that the pilot can at all times monitor the situation.

5. Performance

The single criterion best illustrating the potentialities of a system is its practical performance. Over 4000 safe and successful automatic landings have been made at B.L.E.U. in both propeller and jet aircraft in all sorts of weathers, varying from fogs in which the visibility was as low as 100 ft to winds with components of 25 knots up, down or across the runway, and from very calm to highly turbulent conditions.

An example of the consistency of the system is shown by a series of about 100 landings conducted with a Canberra jet aircraft without resetting any of the electronic equipment. The landings covered weather conditions giving headwind components of 25 knots, tailwind components of 15 knots and cross-wind components of ± 15 knots, some of the conditions being very turbulent. The landings were fully recorded, and an analysis of the figures at touch-down is given in Table 1.

Table 1

	MEAN	R.M.S.
Rate of descent, ft/sec	2	0·9
Longitudinal scatter along runway, relative to glide path origin, ft	470	220
Lateral scatter relative to runway centre-line, ft	10	10
Heading error, relative to runway heading, degrees	½	½

The lateral and heading accuracies need no comment, but the significance of the longitudinal figures may not be so obvious. The longitudinal scatter is, in fact, about half that obtained with average pilots making manual landings in good visibility. This scatter could be further reduced by increasing the mean rate of descent at touch-down, but the chosen figure of 2 ft per second gives a gentle, though positive, landing.

6. Radio Aids

The components of the system just described can be divided under two headings, namely radio aids and autopilot. The autopilot is based on one designed for automatic approach, the chief modification being additions to the computing circuits,[3] but no details will be given here.

6.1. *I.L.S.*

This is an internationally accepted and widely used approach aid, and is fully described elsewhere.[5,6] Briefly, it has two elements, the localizer (azimuth guidance) and glide path (vertical guidance). The localizer is situated at the end of the runway remote from the touch-down point, and generates two overlapping beams in the horizontal plane, the beams being coded by different modulation frequencies and adjusted to give equal signal strengths on the runway centre-line. The carrier frequency is in the band 108 to 112 Mc/s. The glide path is situated close to the touch-down area, and generates two overlapping beams in the vertical plane, the beams being coded by different modulation frequencies and adjusted to give equal strengths on the required approach path, normally about 3 deg to the horizontal. The carrier frequency is in the band 328 to 335 Mc/s. The aircraft must carry two receivers, one for the localizer and one for the glide path frequency, although the localizer receiver can also be used for v.h.f. communications and for v.o.r.

The operating tolerances for the ground equipment are laid down by I.C.A.O. agreement, and are not, as they stand, acceptable for landing. For instance, the deviation of the localizer centre-line from the runway centre-line may be up to $\frac{1}{3}$ deg, and since the localizer aerial may be 10 000 ft from touch-down area, this allows a lateral error of 50 to 60 ft at touch-down, an unacceptably large figure. As for the airborne receivers, certifying authorities and manufacturers have not attempted to meet an accuracy and stability required for landing because this was not the intended use of the equipment. Because of these various limitations, it proved essential to introduce a more precise azimuth aid for landing, the one used so far being Magnetic Leader Cable.

Despite the above limitations of the basic i.l.s. system for landing, there are such obvious advantages in using it to touch-down that the possibility of improving it to the required standard was investigated. At first it appeared practicable because, in modern equipments, localizer beams have been improved in quality by narrowing their widths, and they are more stable than called for in the I.C.A.O. specification to the extent that landing requirements are almost met. More recently, however, a limitation has been found that is believed to be fundamental to the system in its present form, and arises as follows.

With a 10 000 ft long runway (and this is not excessive these days) the localizer will be about 11 000 ft from the touch-down area. An aircraft approaching to land, and at a height of say 200 ft, will be a further 4000 ft from the localizer, and so from 200 ft to touch-down the landing aircraft is below an elevation of 0·8 deg relative to the localizer. With typical localizer effective aerial heights of 6 to 8 ft, the peak of the main vertical lobe is at an elevation of between 20 and 30 deg, so during a landing, the aircraft is on the skirt of this lobe. It has in fact been found that the power reflected from an over-flying aircraft can be so large relative to the direct signal strength at the landing aircraft that the apparent localizer centre-line can be moved right off the runway. The magnitude and persistence of this effect depends on the position, speed and course of the interfering aircraft, but the fact that so large an effect can happen at all and so generate a dangerous manoeuvre when the landing aircraft is close to the ground is unacceptable for reasons of safety. It continues, therefore, to be necessary to substitute an alternative form of guidance for the landing phase.

The glide path does not suffer from this interference problem because the aerial is sited much nearer the touch-down point with the result that the landing aircraft is never close to the edges of its beams. It does not, however, give sufficient information for the flare manoeuvre, so here again it proved necessary to introduce a new guidance element for the landing phase, in this case the f.m. radio altimeter.

It should be emphasized that the criticisms of i.l.s. made above are as a landing aid for which, despite its name, it was never designed. The improvements which have been, and are still being, introduced into equipment and performance are most welcome in its role as an approach aid.

6.2. *Magnetic Leader Cable*

The idea of using the magnetic field from a current-carrying conductor for guidance is not new, and was used operationally in the first World War for guiding ships in narrow channels. More recently systems have been suggested to give three-dimensional guidance information for landing aircraft.[7] These have turned out to be difficult to engineer to give the required height accuracy, but have attractions if limited to azimuth information.

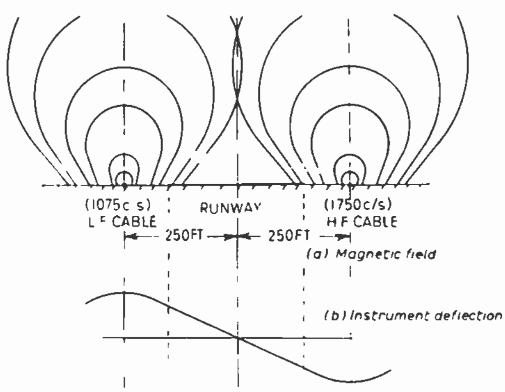


Fig. 2. Leader cable schematic.

The requirements for a landing azimuth guidance system, which leader cable can readily provide, are as follows:—

(i) Coverage must extend to at least 5000 ft into the undershoot area (where the aircraft will be at a height of 300 ft) and may be required down the length of the runway;

(ii) Lateral coverage must be greater than the runway width, and at maximum range must be sufficient to allow for the maximum likely error during the approach phase. A figure of 250 ft each side of the runway centre-line is typical;

(iii) The runway geometric centre-line should be defined to within 5 ft. A change of lateral position of 1 ft should be detectable, and a linear rate term should be obtainable from the output;

(iv) The lateral sensitivity should change as little as possible with height and range;

(v) The equipment should be as simple and reliable as possible, the airborne equipment being also small and light.

Although a single-wire system is possible, a twin-wire system was chosen for the present application for the following reasons:[11]

(i) Whereas the magnetic lines of force associated with a cable in free space carrying current are circles concentric with the cable, this is no longer so when the cable is in or close to the ground. The ground has the effect of "sucking-in" the lines of force (Fig. 2(a)). This limits the unambiguous lateral coverage of the single wire system to well below the 500 ft required, whereas the twin-wire system has no such limitation.

(ii) The single-wire system needs a stabilized search coil in the aircraft, whereas the twin-wire system does not require this complexity.

(iii) The single-wire system approximates to an angular system, so that the output sensitivity varies inversely with height, whereas the twin-wire system has a more nearly constant output sensitivity.

(iv) A single wire, by virtue of the fact that it must be laid down the centre of the runway, may complicate installation.

The system operates as follows: Two cables are laid symmetrically about the runway centre-line and supplied with constant currents of 4 amps, one being at a frequency of 1075 c/s and the other at 1760 c/s. The cables extend 5000 ft into the undershoot area and along the runway for a distance of at least 500 ft past the glide path origin, and perhaps along the full length of the runway. Typical magnetic fields associated with the currents are shown in Fig. 2(a). As long as the cables are symmetrically installed, the plane of the runway centre-line is defined by equality of the two magnetic fields. The airborne receiver is designed to measure the difference between the fields, and therefore gives an output of the form shown in Fig. 2(b). It will be seen that unambiguous information is available anywhere between the cables.

The cable used consists of a 7/0·044 in. conductor, polythene insulated and p.v.c. sheathed, which is widely used as an airfield lighting cable. It can be buried, laid on the ground or mounted on poles, but within the confines of the airfield it is usually buried because of obstruction hazard. Outside the airfield the installation depends on the topography, but it must always be kept symmetrical in the horizontal plane. In the vertical plane considerable tolerance can be allowed, but the cables should not fall more than 80 ft below the runway level in order to ensure there is always sufficient field strength at the aircraft.

A diagram of the system is given in Fig. 3. Each cable is fed at the sending end from its own alternator, one pole of which is earthed. The far end of the cable is also earthed through a capacitive terminating unit. The two alternators, each with an output rating of 1 kVA, are driven on a common shaft by a three-phase 50-c/s motor from the mains supply. The field of each alternator is controlled by a conventional electronic regulator to maintain the cable current constant. Taking for example the low-frequency generator, the output from a current transformer in the cable supply is rectified (D3 and D4) and compared with the voltage of a stable gas tube. The difference is amplified (V3) and applied to valves (V1 and V2) in series with the alternator field. By this means the output current is kept constant to 0·25%, equivalent to a change of centre-line of 3·5 in.

The control circuits, together with monitoring devices, are housed in a console, and this together with the motor alternator are shown in Fig. 4. The total power consumption is 3 kW. Monitoring circuits give an alarm if (i) the currents at the sending end, as measured by instrument current transformers, are out of balance by more than 2·5%, equivalent to a centre-line shift of 3 ft and (ii) currents at the terminating ends, as measured by rectifiers connected across the terminating impedances (the resulting direct currents being sent back along the cables themselves), depart by more than $\pm 10\%$ from the specified 4 amps. Routine control is from a small remote panel, normally housed in the Air Traffic Control Tower, which contains an on/off switch and the alarm just mentioned.

The frequency of the supplies to the two cables depends on the stability of the mains supply frequency, the statutory tolerance of which is $+1\frac{1}{2}\%-2\%$. The

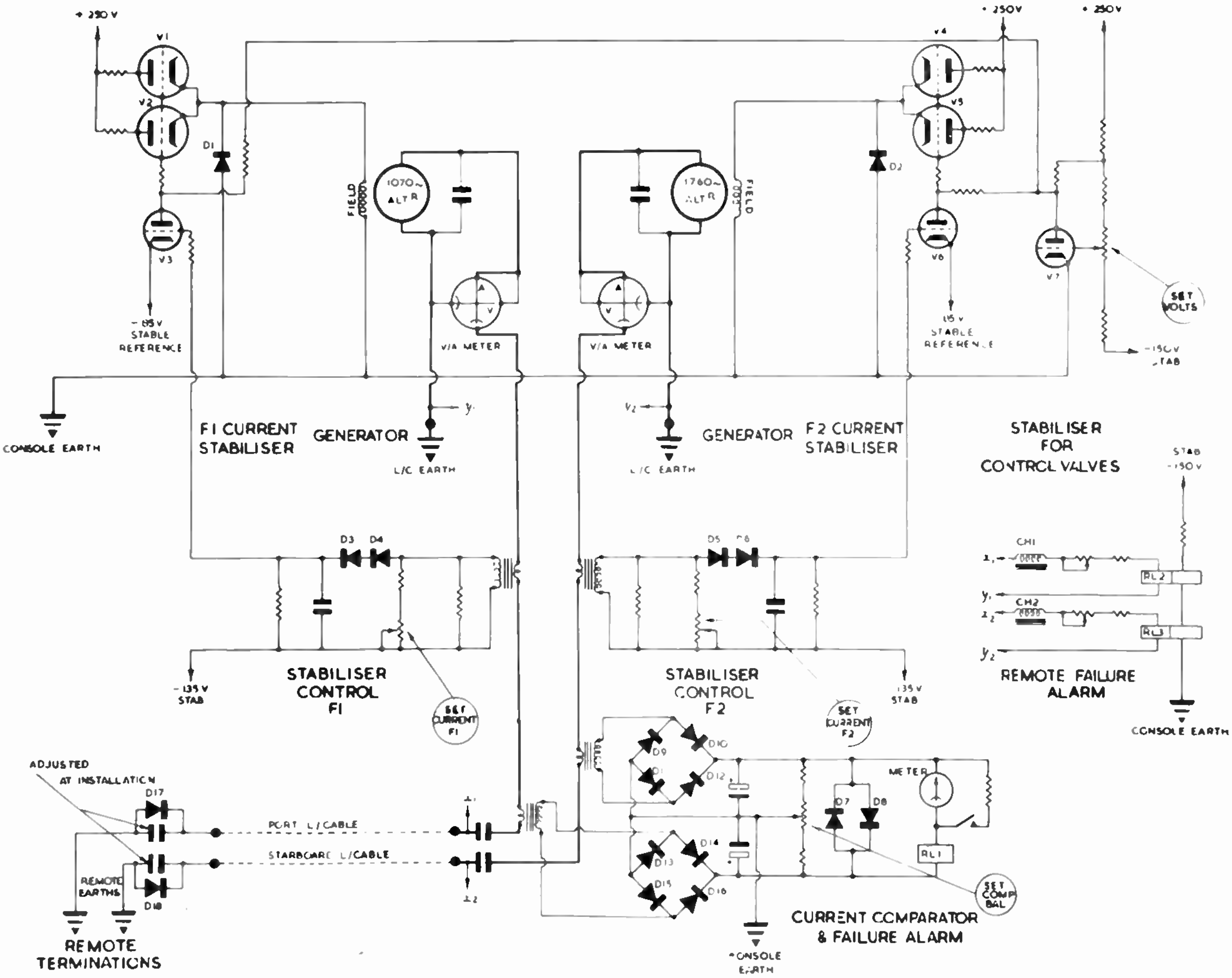


Fig. 3. Leader cable ground installation circuit diagram.

overall system, however, is designed to accommodate a $\pm 4\%$ frequency variation. At initial installation, the termination of each cable, which consists of a number of capacitors that can be connected in parallel, is adjusted to make the current at the sending and terminating ends equal. With typical lengths of cable and with the two frequencies used, the difference between the currents at any other point along the cable does not then exceed $0{\cdot}75\%$ corresponding to a shift in centre-line of 1 ft.

There can be mutual interference problems between leader cables and other cables, although the forms of interference are rather different. For instance, telephone cables can pick up interference from the audio notes from the leader cables, but this can be reduced to a tolerable level by normal balancing techniques. On the other hand significant currents at the leader cable frequency generated by other sources are unlikely, and there is little chance of interference with leader cable from this cause. However, serious field distortions have been found to be caused by earth wires or by screened cables, the screening of which is in contact with the earth. Earth wires, which are prevalent in runway lighting installations, are often run from one lighting fitting to another over considerable distances, the lighting fittings being earthed by being in contact with the ground. This is shown diagrammatically in Fig. 5(a). The leader cable induces currents in the earth circuits, and these can distort the leader cable pattern to an unacceptable extent. This trouble can be cured by breaking the earthing wire into lengths of not more than 400 ft and earthing each length near to its centre, as in Fig. 5(b). Comparable treatment has cured other interference of this kind that has so far been met. Luckily the distorting field falls off very rapidly with distance, and so, taking into account that the landing aircraft is only close to the ground just before touch-down, it is found that the critical area for interference of this kind is confined to the runway area and about 500 ft into the undershoot. Even though remedial action may be necessary in this area, at least it is totally contained within the airfield, so that all services are under the control of the airfield authority.

The airborne receiver consists of three units, namely the loop assembly, the amplifier unit and the control unit. They are shown in Fig. 6 and their total weight is 20 lb.

The loop assembly consists of two loops mounted at right angles, each tuned to one of the leader cable frequencies, and a miniature 400 c/s 3-phase motor that rotates them at a speed of 1000 rev/min about an axis parallel to the aircraft's longitudinal axis. The loops must be mounted at least 6 inches clear of the aircraft's metallic skin, and close to the centre-line. The rotation applies a 30 c/s modulation to the received signals, but use is made later in the receiver only of the peak value of this signal. By this means only the modulus of each magnetic field is measured, and effects of the direction of the field and of the aircraft attitude are eliminated.

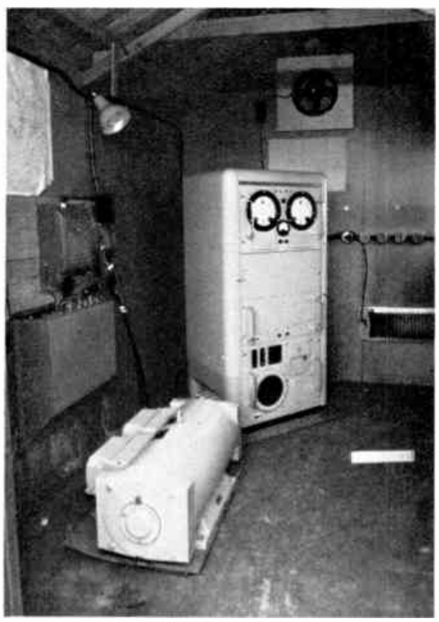


Fig. 4. The control console and motor alternator.

Referring to Fig. 7, the signals from the two loops are mixed at the input to the amplifier, (V1 to V3) which consists of three conventional stages with a.g.c. applied to the first two, has a flat frequency characteristic to 3 kc/s, and has a maximum gain preset to 90 dB, which is sufficient to allow operation to a height of 450 ft. At the output of the amplifier the two frequencies are separated by the two M-derived filters. These, taken in conjunction with the tuned input coils, have an overall response shown in Fig. 8, the separation between the two frequencies being not less than 30 dB. This is adequate since in normal operation the difference in level between the signals at the two frequencies does not exceed 20 dB. The output from each filter is rectified (D1 and D2) with a smoothing time-constant of $\frac{1}{4}$ sec, and then fed to cathode followers (V4, V5). The final output from the receiver is taken from between these two cathode followers, and is fed to the pilot's instrument display and to the autopilot.

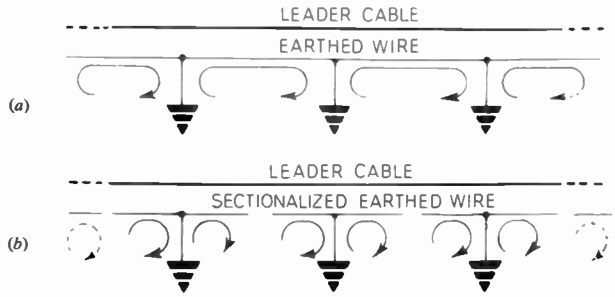


Fig. 5. Leader cable installation; effect of induced current in earthed conductors.

The only unusual feature of the receiver is that its a.g.c. signal is derived from the smaller of the two leader cable signals. This is done for three reasons:

(i) it provides a safety feature against receiver malfunction, in the following way. If a fault should stop one of the two channels from functioning, so that one signal disappears, the a.g.c. will not operate and so leader cable signal will never be switched into the autopilot. If a.g.c. were taken off the maximum signal, then leader cable would be selected, even if one channel were faulty, and the aircraft would be guided off the runway;

(ii) the output signal stays linear over a greater range of lateral displacement;

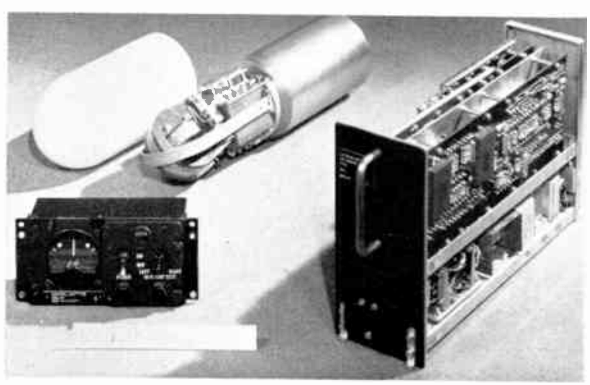


Fig. 6. The airborne receiver, comprising (*left to right*)—control unit, loop assembly and the amplifier unit.

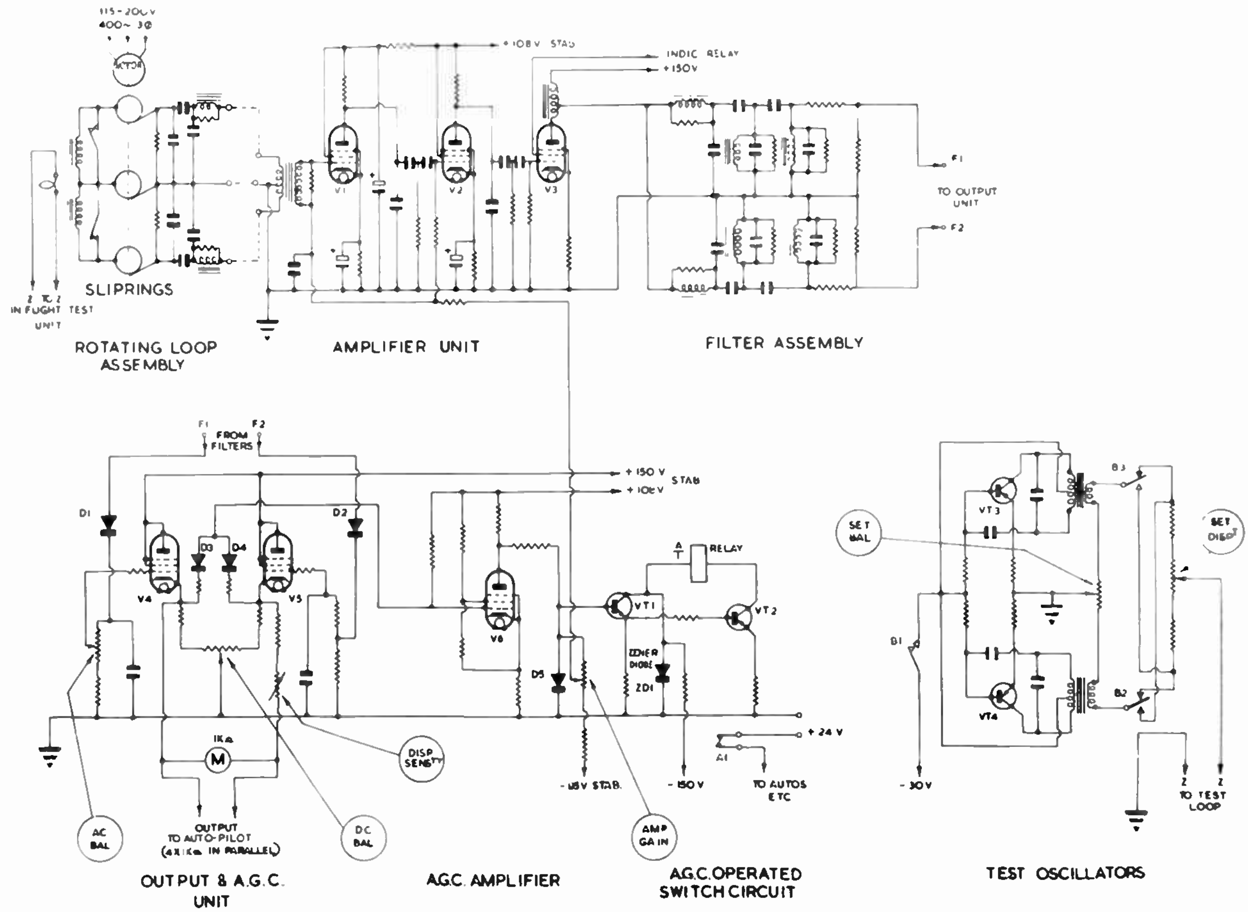


Fig. 7. Leader cable airborne receiver circuit diagram.

(iii) the sensitivity of the output signal stays more constant with change in height.

The a.g.c. circuit functions in the following manner. The diodes D3 and D4 have their positive ends commoned and connected via a resistor to h.t., while their negative ends are individually connected to each cathode follower. The common ends of D3 and D4 therefore stay at the same potential as the lower of the two cathode follower voltages, and in this way the smaller of the two leader cable signals is selected. This voltage is applied to the grid of V6, the cathode of which is biased positive, so introducing some delay to the a.g.c. An amplified signal from the anode of V6 is applied as a.g.c. voltage to the grids of the amplifier valves, V1 and V2. Diode D5 and the amplifier gain control determine the least negative signal that can be applied to the grids of V1 and V2, and therefore the maximum gain of the amplifier.

The output from V6 is also applied to the transistor amplifier VT1 and VT2 which drives a relay. It is this relay that switches the leader cable signal into the autopilot, in place of the i.l.s. localizer signal, when the leader cable receiver a.g.c. operates.

There is an in-flight test facility built into the receiver so that its performance can be checked immediately prior to automatic landing. Two self oscillators, VT3 and VT4, can be energized, each tuned to one of the leader cable frequencies, and the output currents passed through a single-turn coil mounted in the

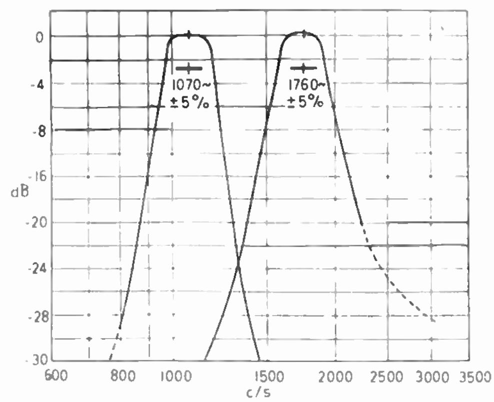


Fig. 8. Overall response curve for the leader cable airborne equipment.

search-coil housing. In the small control unit is mounted a centre-zero meter, and the test facility can be selected by a switch on the unit giving either a predetermined deflection left or right. The test signal is of sufficient strength to operate the a.g.c. and this is indicated on a "doll's-eye" type indicator. If all these register correctly the complete airborne receiver is functioning satisfactorily. Also mounted on the control unit is the main on/off switch. A self-contained test set is available for ground testing the airborne receiver.

The leader cable system meets the requirements stated at the beginning of this section. Both the ground and airborne equipment are simple and robust, and the fact that their performance is acceptable is perhaps best illustrated by the results of lateral errors quoted for the overall system in Section 5. The only disadvantages appear to be that some treatment of other installations in the runway area may be necessary to reduce interference, and that laying rights will have to be obtained outside the limits of the airfield boundary. The length of 5000 ft required in the undershoot is set by the time taken for an aircraft to perform the S-turn manoeuvre required to reduce its lateral error from the runway centre-line, and taking into account the increasing approach speeds of aircraft, it seems unlikely that it will be possible to reduce this length significantly.

6.3. *F.M. Radio Altimeter*

As has already been mentioned, a satisfactory technique for executing the flare manoeuvre has been found to consist of controlling the aircraft below about 60 ft by making the rate of descent proportional to height, the resulting flight path being exponential. It is also convenient to have accurate height up to about 500 ft in order to initiate various phases of the automatic landing and so that the pilot may have as a monitor a more reliable height above ground than from his pressure altimeter. These requirements for both accurate height and rate of change of height demand rather stringent characteristics from the altimeter.

Measuring height relative to the ground calls for an electronic altimeter, and at the low altitudes required an f.m. rather than a pulse technique is more appropriate. Frequency-modulated altimeters have been available since the 1930s, but special development was necessary to provide the characteristics required for automatic landing.[12]

These are:

(i) it must function between a height of 2 ft (the lowest height the aerials are likely to reach when mounted in a stationary aircraft) and a few hundred feet, typically 500 ft.

(ii) the output must be linear, and any smoothing time-constant in the output must not exceed 0·25 sec, so that the output can be differentiated to give a rate of change of height without this signal being too much delayed.

(iii) the zero error must not under any circumstances exceed 2 ft while the additional error at increasing height must not exceed 3% of the true height.

The mode of operation of f.m. altimeters has been thoroughly covered in the literature.[12,13,14,15] However, the following very brief description may serve as a reminder of the main features.

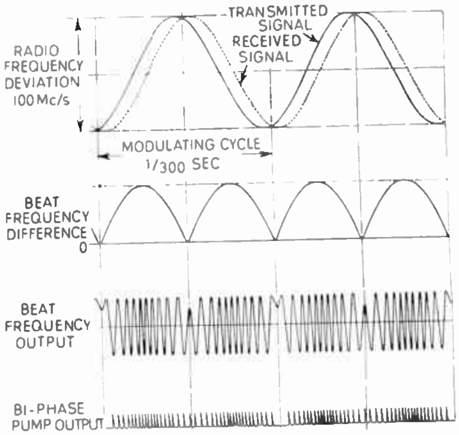


Fig. 9. Determination of beat frequency in the radio altimeter.

The f.m. altimeter functions by virtue of the received signal reflected from the ground being mixed with the signal then being transmitted, the beat note thereby generated being proportional to height. This is illustrated in Fig. 9. If the modulation is not linear, the beat note will vary in frequency during the modulation period, but the number of cycles in the beat note during this period depends only on the swept band. If then a cycle-counting technique is adopted, it is not necessary accurately to maintain linearity of modulation, or even any given modulation waveform, as long as the swept band is kept constant. Cycle counting is accomplished in the present altimeter by generating a pulse at each zero crossing of the beat note, and then integrating this train of pulses.

This technique eases requirements on the modulation waveform, but it brings in a problem of its own. Imagine that the altimeter is suspended over a perfectly conducting earth and that it is being raised very slowly but at a constant rate. Because of the changing r.f. phase between the transmitted and received signal, the beat note pulses, observed within the frame of a modulation half-period, will move across, and one

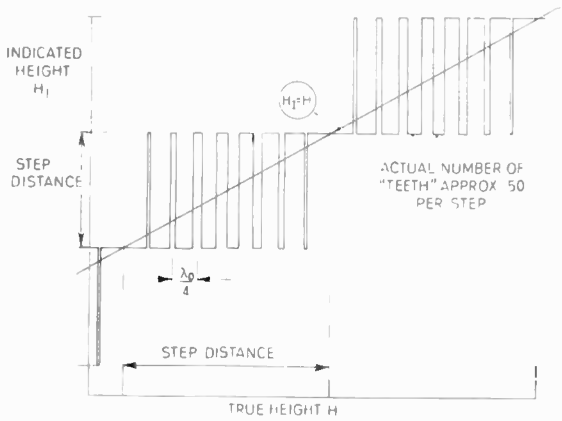


Fig. 10. The step effect experienced with the radio altimeter.

pulse may drop off the end of the frame before another one appears at the other end. Thus the "count" may vary in steps by one pulse, giving a "step distance" in the height output signal. This step distance is independent of carrier frequency, but can be shown to be equal to one quarter of the wavelength equivalent to the swept frequency band.

Returning to the concept of the altimeter being very slowly raised in height, one complete cycle of a pulse dropping off one end of the beat note train and then reappearing at the beginning will be generated each time the height changes by $\frac{1}{4}$ wavelength at the carrier frequency, while at the same time, when the altimeter is raised over many such quarter wavelengths, the pulses will get closer together, i.e. the beat note will increase in frequency. The nett effect is that the theoretical height output, in these ideal conditions, is as shown in Fig. 10, in which λ_0 is the wavelength of the carrier.

Two things now become apparent. The larger the swept band the smaller will be the amplitude of the steps in the height output, and the higher the carrier frequency the smaller will be the changes in true height that will cause a step in the height output. Both these features are important for a flare-out altimeter. The need for a small step distance, in order to achieve high accuracy, is obvious. A high carrier frequency, i.e. a small λ_0, ensures that during normal use the steps in the output height occur very rapidly, because small true height fluctuations are caused by perturbations to the aircraft flight path, by small undulations even in so-called level ground and because during the flare-out manoeuvre the aircraft is constantly changing height. These rapid output height fluctuations can then be sufficiently smoothed in the allowable output time-constant of 0·25 sec for the output to be differentiated to give a usable rate-of-change-of-height term. Fortunately, because a wide swept band also calls for a high carrier frequency in order (a) not to require too high a percentage bandwidth frequency allocation and (b) to ease the design of the r.f. circuits, the technical solutions to both these requirements go hand in hand.

For the present application it was decided to operate in the allocated band of 4200 to 4400 Mc/s. This

Fig. 11. The radio altimeter equipment.

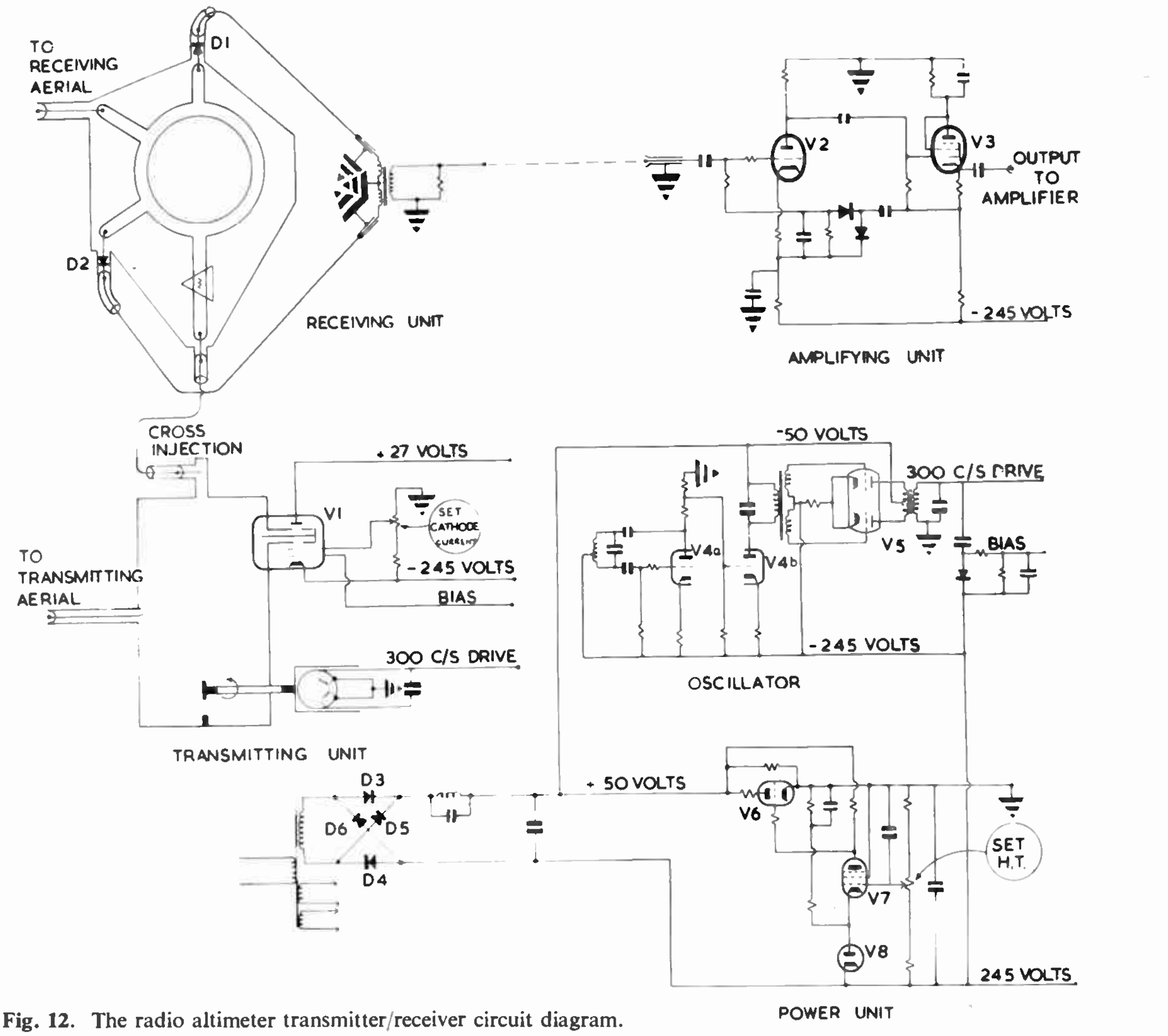


Fig. 12. The radio altimeter transmitter/receiver circuit diagram.

means that the steps shown in Fig. 10 occur for a change of height of only about $\frac{3}{4}$ in. Also it allows a swept band of 100 Mc/s to be used, giving a step distance of $2\frac{1}{2}$ ft or a zero accuracy of $\pm 1\frac{1}{4}$ ft. However, it is found in practice that owing to the smoothing of the output steps described in the previous paragraph, the accuracy of the measured height at heights below 50 ft is about 1 ft. At the higher heights the accuracy is not worse than 3%, as required by the specification.

The latest version of the altimeter, considerably smaller than a previous version, is shown in Fig. 11. It should be mentioned here that it was not developed entirely as an autolanding altimeter, but also as a so-called general purpose altimeter, and it has two height ranges of 0 to 5000 ft and 0 to 500 ft. It weighs only 25 lb, but if it were produced as a single range instrument for automatic landing only, it could be reduced in size and weight yet again.

The altimeter consists of separate transmitter and receiver aerial horns, an r.f. delay cable assembly, a transmitter and receiver-head-amplifier unit, a receiver (counting unit) and a control box.

The aerial horns have an aperture of 5 in. by 4 in. and a polar diagram of 34 deg included angle between 3 dB points fore and aft and 40 deg at right angles. So that true vertical height is measured during normal aircraft attitude changes, the horns must be mounted so that the polar diagrams are within 5 deg of the vertical. They must also be separated by at least 15 in. in order to reduce the direct coupling between them to less than 60 dB.

A circuit diagram of the transmitter and receiver-head-amplifier unit is shown in Fig. 12. The transmitter consists of a cavity excited by a Heil tube (V1). This has the advantages of being rugged and of permitting frequency modulation with little attendant amplitude modulation. The carrier frequency is in the band 4200 to 4400 Mc/s and the swept band is 100 Mc/s on the 500 ft range and 10 Mc/s on the 5000 ft range. Modulation is achieved mechanically by rotating a paddle wheel inside the cavity. The swept band is adjusted by varying the insertion of a post, while range is changed by altering the insertion of the whole paddle assembly. The modulation frequency is 300 c/s, and the motor is driven from a special stable 300 c/s oscillator (V4 and V5). This is because, for simplicity, the counting is not done during each modulation cycle but instead over a fixed time, so that accuracy of height output is dependent on the constancy of the modulation period. The modulation waveform approximates to a sine wave. Output to the transmitter horn is from a fixed coupling loop in the cavity via an r.f. concentric cable, and the radiated power is 750 mW. There is another low level output from an adjustable probe to inject a reference signal into the receiver.

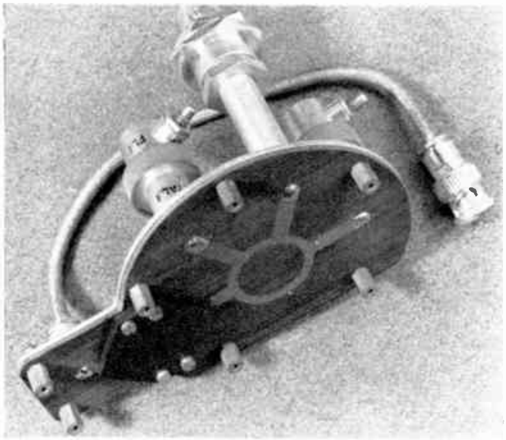


Fig. 13. The stripline mixer assembly.

The receiver horn is connected to the receiver by an r.f. concentric cable. Here it is fed direct to a balanced mixer consisting of a rat-race, the "local-oscillator" signal being the above-mentioned injection from the transmitter. The novel feature of the mixer is that it is made of stripline instead of waveguide, in order to reduce weight, and a photograph is shown in Fig. 13. The output beat-note is amplified in a low-noise valve preamplifier (V2 and V3) before being passed to the main amplifier and counting circuits in another unit.

The amplifier and counting circuits are entirely transistorized, and a circuit diagram is shown in Fig. 14. With the transmitter parameters chosen, the beat note turns out to be about 120 c/s per foot on the 0 to 500 ft range. The amplifier, therefore, has to handle frequencies from 300 c/s (the modulation frequency) to about 120 kc/s, the approximation being introduced because the beat note is in fact varying because of the non-linear modulation waveform. The upper frequency is set equivalent to 1000 ft instead of the "maximum" height of 500 ft because there is a requirement that the height indicator meter should remain hard against its upper stop to a height of at least twice the scale maximum.

There is always a direct leakage signal from transmitter to receiver horn, although this is limited to at least 60 dB below transmitter power by the spacing between the horns. However as height, and therefore beat note, increases, the signal strength of the wanted signal reflected from the ground obviously reduces, and so, at the higher heights, some further discrimination between wanted and leakage signals is essential. This is introduced by giving the amplifier (VT1 to VT8) a rising characteristic, so that the gain is larger at the higher frequencies. Unfortunately at very low heights an effect appears which requires the opposite characteristic. At heights below about 100 ft, where signal strength is very strong, it is possible to get a "double-bounce" between aircraft and ground, so that an unwanted signal of considerable magnitude can appear at a frequency of twice the wanted signal. If the amplifier has a rapidly rising characteristic, then this signal receives preferential treatment and can introduce large errors in indicated height. To overcome this difficulty, a variable resistance element (v.r.e.) is introduced into the receiver. A transistor (VT2) is made the variable resistance element of a filter and is fed from the output height signal. At a high height, the output counted signal is large, and the v.r.e. transistor is biased to a low resistance state so that the filter cuts off below a fairly high frequency. As the height is reduced, the output counted signal is reduced, and the v.r.e. transistor resistance increases, so that the filter now cuts off below a lower frequency. The effect is shown in Fig. 15, where it will be seen that the characteristic is automatically adjusted so that there is a rapid cut off a little below the frequency being measured, while above this frequency the characteristic is only slightly rising.

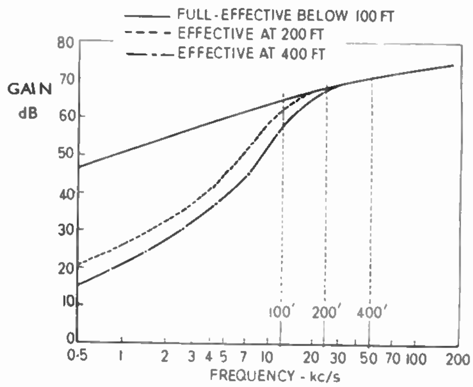


Fig. 15. Amplifier characteristics of the radio al·imeter.

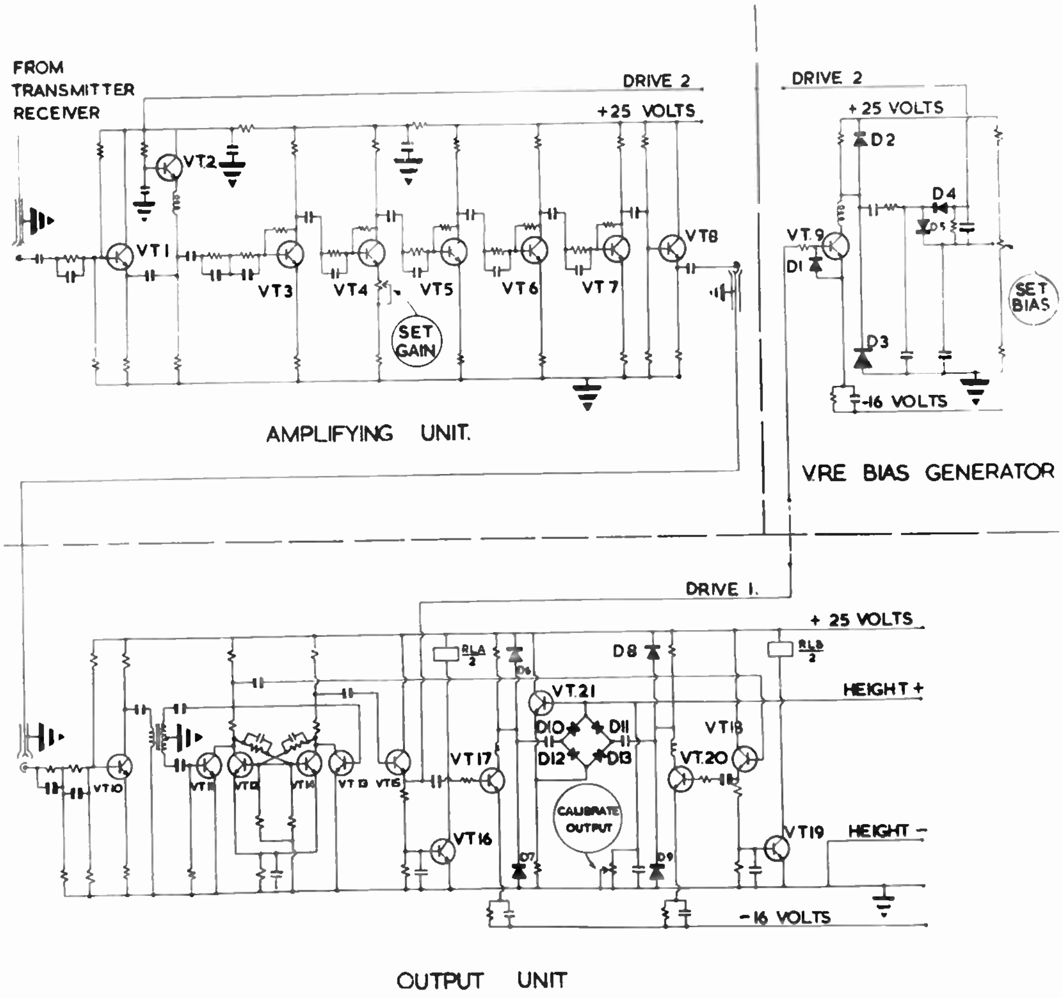


Fig. 14. The radio altimeter amplifier circuit diagram.

Referring to Fig. 14, the main amplifier VT1 to VT8 has the characteristic already referred to. The output is converted to push-pull in the output transformer of VT10, which drives, through trigger stages VT11 and VT13, the Eccles-Jordan relay VT12 and VT14. This converts the beat-note signal to a square-wave with constant slope at zero crossings. Each square-wave output is fed through an emitter-follower, VT15 and VT18, to driver stages VT17 and VT20 of a full-wave push-pull rectifier, or diode pump, D10 to D13, which constitutes the actual counting circuit. Diodes D6 to D9 are limiters to ensure that the input amplitude to the counting circuit is always constant. VT21 is a "boot strap" emitter-follower which enables the diode pump to give a linear, instead of an exponential, output current against input frequency.

As has already been mentioned, VT9 and diodes D4 and D5 form a separate counter, full-wave but this time not push-pull because a lower current output is required, which drives the v.r.e. stage VT2.

Transistors VT16 and VT19 and their associated relays are monitor circuits. If the square-wave departs markedly from a unity mark/space ratio, one or other of these relays will operate. Their contacts are connected in series so that if either operates a warning flag appears on the altimeter meter. Departure from unity mark/space ratio shows that the circuit is operating on noise instead of a genuine signal, and the flag warns the pilot to disregard any output indication.

An in-flight test feature is introduced so that the correct functioning of the whole altimeter can be checked immediately prior to it being used for automatic landing. By selecting "test" on a switch on the control unit, a probe is introduced, by relay action, into the throat of each horn. These probes are connected by an accurately known length of r.f. concentric cable, and in this test condition the transmitter power is diverted along the cable into the receiver, so that the output reads the electrical length of the cable. As long as the known reading is obtained at the output, the equipment must be functioning satisfactorily. It even checks the power output, because the attenuation down the cable is equivalent to that along the normal path from average ground.

The output from the altimeter is fed to an indicator mounted on the flight panel, connected in series with the input to the autopilot. The differentiating circuit to give rate of change of height is part of the autopilot computer.

A small control box contains the on/off switch and the in-flight test switch.

A self-contained test set is available for ground testing the altimeter.

This altimeter meets the requirements stipulated at the beginning of the section and once again the fact that its performance is acceptable is probably best illustrated by the longitudinal scatter of touch-down point obtained from the overall system which was quoted in Section 5. This scatter is about half that obtained from manual landings in good visibility.

7. Application to Civil Aviation

The system so far described has already been adopted by the R.A.F., but the civil operator, while he has been aware of the development, has only become actively interested during the last two years and it is over this period that his special problems have been studied. In the main they reduce to two: safety, which is the concern of any blind landing system, and choice of an agreed guidance aid, which to some extent is a problem of this particular automatic system.

7.1. *Safety*

It will be evident that the compromise chosen between such factors as size, weight, performance, reliability etc may well be different for military and civil versions of an equipment. The automatic landing system so far described has the right compromise for some military uses, but for civil application an even higher reliability is required, at the expense if necessary of some weight penalty. The aim is to ensure that a civil blind landing system does not cause a fatal accident more frequently than occurs during manual landings, which dictates that the automatic system must not cause a fatal accident more frequently than 1 in 10^7 landings.

The safety problem can be divided into two stages. The immediate and essential aim is to ensure just before an automatic landing is attempted that sufficient safeguards are available to make the final landing adequately safe. During this stage diversions will still be made in bad visibility if *en route* failures have reduced the safeguards available at landing below an acceptable level. The longer term aim is to establish scheduled automatic landings, that is to commit the aircraft to an automatic landing before the flight begins.

As far as the first stage is concerned, automatic systems are already available that give a fault warning and allow a pilot to take over control with little or no deviation from the wanted flight path, and these systems are adequate for use on the approach down to a height of perhaps 200 ft, because if a fault should happen in bad visibility the pilot can still take over, overshoot and divert to another airfield where the visibility is suitable for a manual landing. Below a height of about 200 ft, though, it is imprudent in bad visibility to hand back the aircraft to the pilot if the autopilot develops a fault, because, even if he had adequate guidance information on his flight instruments, he would be hard pressed to take over the aircraft and successfully complete an instrument landing in the short time available. This makes it essential to ensure that the automatic system itself meets a reliability figure of 1 in 10^7 over the last 30 seconds or so of the landing. The only way of achieving this appears to be through redundancy, but the exact form and amount of redundancy remains open to argument. Current opinion favours between two and four similar, if not identical, sets of equipment, a comparator technique being used to eliminate faulty elements.

The longer term aim is even more difficult and this is to ensure at take-off that the equipment will be functioning satisfactorily to meet the required safety figure at landing after a flight of several hours. If this could be achieved the dividends are great, because diversion fuel could be dispensed with, thereby increasing by a large factor the payload of the aircraft and giving the opportunity of significantly reducing fares. The only thing certain about this problem is its magnitude. Whether it will be solved by yet more redundancy or by some form of in-flight monitoring on a unit basis is by no means clear.

Whatever overall system philosophies finally prevail, it seems certain that the basic system should be kept as simple as possible, perhaps even at the expense of sacrificing improvement in performance, and that individual units should have a reliability as high as the engineering art can provide. Simplification should also reduce weight and volume, of particular importance where redundancy is being considered.

It is probable that the best way of performing the flare manoeuvre will continue to be by making direct use of height above ground rather than computing it from range and elevation information. There is therefore strong reason for attempting to simplify height measurement between say 150 ft and ground. One method would be to develop an f.m. radio altimeter specifically for this height range, which would then be smaller and lighter than the existing general purpose model, and might even be reduced to common aerial working, thus simplifying aerial installation. Attention is already being given to this proposal. Another possibility with this type of altimeter is to make use of the absolute test of measuring the electrical length of a cable (used at present only as a spot in-flight test) as a continuous check on the correct functioning of the equipment, so reducing the redundancy otherwise required.

The azimuth system used at present, i.e. leader cable, is inherently simple. The obvious improvements to the airborne equipment would be to remove the need for rotating the loops, if this is possible, and to transistorize the amplifier in order further to reduce its size and weight. It must not be forgotten that the ground equipment must be equally as reliable as the airborne, so that the chance of it failing during the time an aircraft is landing must also be less than one in 10^7. In the case of leader cable it should be possible to achieve this by normal power engineering methods. More will, however, be said about leader cable in the next section.

7.2. *Choice of Guidance System*

It has already been suggested that the simplest form of guidance for the landing flare-out manoeuvre may well be direct measurement of height above ground. This has the added advantage that the equipment is entirely airborne, so that the choice of the type of equipment to be carried lies only with the airline. Unfortunately the same is not true for the azimuth guidance, where some ground equipment is essential, and so a system must be chosen that is likely to be acceptable to the authorities, mostly government agencies, who will have to install it.

It may well be that ground aids for landing will, like the safety problem, pass through two stages. In the first place the landing aid will be living for some time alongside the i.l.s. approach aid (which is protected until 1975), and a short-range azimuth aid may be all that is required that takes over during the final stage when the i.l.s. localizer can no longer be used. On a longer term basis there will undoubtedly be a replacement for i.l.s. which will need characteristics to accommodate both approach and landings and must form part of the overall Air Traffic Control System to be used in the following decades. It therefore appears that before system design work can proceed on this long-term project, some statement is required of the operational characteristics needed.

Returning to the more immediate problem of a shorter-term aid, leader cable is the only one whose performance has been rigorously tested and found adequate. Its simplicity also has the strongest of attractions. However, the ground installation suffers from the disadvantage that land outside the airfield boundary must be acquired, since the cable must extend to about 5000 ft from runway threshold and little chance is seen of significantly reducing this length. This conflict must be resolved if there is to be early adoption of the system.

Work is in hand to investigate alternative azimuth guidance systems that would be easier to install while preserving the performance and as much as possible of the simplicity and reliability of leader cable. If such a system is found it will of course take some time to develop to the same proved standard as leader cable. Choice appears to be limited to two main types, either a comparatively high frequency system in which the energy is beamed into the smallest elevation and azimuth angle tolerable for approach and landing or a comparatively low frequency system in which obstacles do not significantly distort the field pattern. Both types of system are being investigated.

8. Conclusions

The guidance elements have been described of an automatic landing system which has a performance adequate for blind landing and a consistency of touch-down better than achieved by the average pilot landing manually in good visibility. There are, however, a number of engineering problems to be solved to achieve the reliability and safety required by civil aviation. These problems are common to any blind landing system and not specific to this particular one.

In addition the problem concerning leader cable must be resolved of whether its good performance and simplicity outweigh problems of installation. Alternative precision azimuth guidance systems are being sought.

9. Acknowledgments

The system described is the result of a team effort. B.L.E.U. has been responsible for the concept and proving of the system, while industry has developed the equipment to production standard. Of the guidance aids described in this paper, the Magnetic Leader Cable ground and airborne equipment was developed by Murphy Radio Ltd. and the F.M. Radio Altimeter by Standard Telephones & Cables Ltd. The opinions expressed are, however, entirely those of the author.

10. References and Bibliography

1. W. J. Charnley, "An Automatic Landing System with Comments on its Civil Application". International Air Transport Association. Report of 11th Annual Technical Conference of I.A.T.A. at Monte Carlo in September, 1958.
2. W. J. Charnley, "Blind landing", *J. Inst. Navigation*, **12**, No. 2, April 1959.
3. W. J. Charnley, "The Work of the Blind Landing Experimental Unit". Seventh Anglo-American Aeronautical Conference, New York, October 1959.

4. D. Lean, "A Flight Study of the Problem of Correcting Lateral Errors during the Final Visual Approach". Appendix A of the Report of the Ninth Technical Conference of I.A.T.A., San Remo, pp. 237–263, May 1956. Doc. Gen./1650.

5. International Standards and Recommended Practices, Aeronautical Telecommunications. Annex 10 to the Convention of International Civil Aviation. Published by International Civil Aviation Organization.

6. M. Birchall, "C.w. radio aids to approach and landing", *J. Instn Elect. Engrs*, **94**, Part IIIA, p. 943, 1947.

7. E. N. Dingley, Jr., "An instrument landing system", *Communications*, June 1938.

8. J. Blanchard, "Le Balisage des Aerodromes par Cables Enterres", *Rev. Gen. Electricité*, **24**, No. 2, pp. 765–766, 1953.

9. G. G. MacFarlane, "The Magnetic Field of a Leader Cable Suspended above the Earth". T.R.E. Memo, No. 20, 1950.

10. Staff of B.L.E.U., "Automatic Flight Tests using Leader Cable Azimuth Control following an I.L.S. Approach". R.A.E. Tech. Note No. BL.20, June 1950.

11. B. D. W. White, "Leader Cable Azimuth Guidance Systems for Aircraft Landing". R.A.E. Tech. Note No. BL.25, September 1953.

12. M. P. G. Capelli, A. E. Outten and K. E. Bücks, "The application of radio altimeters to aircraft approach and landing", *Proc. Instn Elect. Engrs*, **105B**, pp. 358–64, Supplement No. 9, 1958 (I.E.E. Paper No. 2587R).

13. M. P. G. Capelli, "Radio altimeter", *Trans. Inst. Radio Engrs (Aeronautical and Navigational Electronics)*, **ANE-1**, No. 2, pp. 2–7, June 1954.

14. B. A. Sharpe, "Low-reading absolute altimeters", *J. Instn Elect. Engrs*, **94**, Part IIIA, pp. 1001–11, 1947.

15. P. C. Sandretto, "Electronic Avigation Engineering". (I.T. & T. Corporation, New York, 1958).

16. A. M. A. Majendie, "Some Considerations of Safety in Automatic Flight Control". Report on the first International Congress of Aeronautical Sciences, Madrid 1958.

17. K. Fearnside, "Trends in the instrumental and automatic control of approach and landing", *J. Inst. Navigation*, **12**, pp. 66–83, January 1959.

Manuscript received 11th April 1960 (Paper No. 599).

DISCUSSION

Under the chairmanship of Mr. K. E. Harris, B.Sc. (Member)

Mr. C. N. W. Reece (*Associate Member*): Has the author experienced any field distortion in the leader cable system due to, say, local concentration of earth return currents caused by non-linear earth conductivity? Has he been particularly fortunate at Bedford, or does he consider that such current concentrations would be too low to effect the exterior field?

Considering that reliability is a prime factor of this system, I would have thought it desirable to eliminate the rotating coil method of field resolution with its mechanical parts and slip rings. Has the author tried the static three-coil method of resolving the field and rejected it for some reason?

The author mentioned an auto-throttle coupling between throttle and airspeed during the final approach of the aircraft. It would appear to me that if the coupling was between throttle and angle of attack (or incidence) this would be more sensible, as it would always give the right feedback irrespective of aircraft loading.

Mr. M. Capelli: How is the datum landing level determined for the flare-out? This figure must vary with different types of aircraft because of their different behaviour at "touch-down". Is the datum determined empirically by conducting flight trials on each aircraft type?

The amount of drift "kick-off" is determined by the difference between the aircraft heading and runway bearing. Is the runway bearing set in manually by the pilot? If so, what errors are encountered in this procedure.

Mr. P. F. Mariner: May I say something in support of Mr. Shayler with regard to the uses of radar. Whilst being a radar manufacturer and therefore interested in the uses of radar, I cannot help but feel that there will be considerable difficulties in the way of using radar for this purpose. In order to provide a safety reliability of 1 in 10^7, a time between failures of the order of 80 000 hours is required. I do not think anyone would suggest that one should envisage making a radar with a safety reliability of this order and therefore some degree of redundancy will be required. Redundancy, as we have heard, is provided by duplication, triplication or quadruplication of the equipment. It is difficult enough in modern aircraft to find a siting place for one radar, let alone for four radars and, whilst I have considerable sympathy for this sort of system and have, in fact, proposed such systems myself, I think this solution is an extremely difficult one and one which cannot obviously be solved by techniques existing at the moment.

Mr. B. Williams: I have gained the impression during this discussion that most of the speakers regard i.l.s. as a system of such poor accuracy as to be only useful in an automatic landing system, in azimuth, at distances greater than 1 mile from touch-down, and in elevation at heights greater than 150 ft on the glide path. Whilst this may be true of i.l.s. systems which just comply with I.C.A.O. regulations for an approach aid to visual landings, it is certainly not the case with the present system developed and produced for military and civil use by my company, and used by B.L.E.U. in its automatic landing system.

This system is capable of delineating the centre-line of the runway or its extension into the approach area to an angular accuracy of $\pm$ 0·05 deg and of maintaining this accuracy over long periods of unattended operation. In elevation glide paths of nominal slopes of from 2 to 4 degs can be set up to accuracies of better than $\pm 0{\cdot}05\theta$ (θ is the glide angle) and maintained down to aircraft aerial heights of less than 40 ft. This is well within the range of takeover by the radio altimeter in the B.L.E.U. system and if required removes the gap between 100/150 ft to 60 ft over which possible lack of control has worried previous speakers.

In fact, a large number of completely automatic landings have been made without the use of leader cable and using Pye i.l.s. as the azimuth guidance element with lateral spreads at touch-down of about twice those obtained with leader cable. The elimination of leader cable is not yet operationally acceptable owing to the probability of the occurrence of transient course line distortions due to the re-radiation of the i.l.s. field patterns from other aircraft not under operational control and crossing the runway centre-line precisely within 2 to 3 seconds of the aircraft on an automatic landing being close to touch-down. These effects are being investigated.

The present B.L.E.U. system necessitates the installation of leader cable for some considerable distance outside the boundaries of most airfields and raises problems of way-leave installation, security and maintenance which for civil airports may be insoluble. In the case of London Airport the leader cables would pass over two major roads and I would like to ask Mr. Shayler if he considers the continuous passage of large steel vehicles over the extended run centre-line would produce any significant distortion of the leader cable field patterns which might lead to effects similar to those possible when using i.l.s. alone.

AUTHOR'S REPLY

I can assure Mr. Reece that no effect of non-uniform ground conductivity has been experienced. It is to be expected from theoretical considerations that the return current would be distributed diffusely in the ground, and this has been confirmed experimentally. Consequently fine-grain variations in earth constants should have negligible effect.

When the leader cable receiver was designed some time ago, fixed-coil resolver systems were considered. At the time it was decided that mechanical rotation was simpler and more reliable, and certainly it leads to extremely simple circuits. If the receiver were being re-designed now, resolver systems would certainly be re-explored, but I should not like to pre-judge the issue until I had seen details of the circuit complexity involved in the resolver system.

Tests have been done using incidence instead of airspeed in the automatic speed control, and difficulties were encountered owing to errors and noisiness in the instrument that measured incidence. It is true that if a suitable signal were available, the pilot would be relieved of selecting an approach airspeed. On the other hand, the automatic speed control can be used in flight régimes other than the approach, and here the pilot needs to set specifically his airspeed. The present system therefore provides flexibility.

Mr. Capelli is right in assuming that the height datum is different for each aircraft type. It is one of the parameters which must be correctly determined to adjust the mean rate of descent at touchdown to 2 ft/sec. The value is first found by use of a computer, into which are set the control equations and the aircraft aerodynamic derivatives. In fact this is how all the control constants are explored. Final adjustment is made during type-test flight trials.

The runway heading, required for kicking-off drift, is set into the beam compass manually by the pilot. There is nothing new here, because the same information must be set in for joining the i.l.s. localizer beam during the standard automatic approach. Consequently the pilot is used to making this adjustment prior to the approach phase. His inaccuracy should not be worse than 1 deg.

The question, put by another speaker, of why we do not use well-established precision radar techniques for blind landing is one that is often raised, and I am particularly glad to get Mr. Mariner's support in answering it since he himself is in the radar field. His reply emphasizes a point I have tried to make in my paper, that it is not only performance that is required but also reliability, and that the two must be considered together in deciding on a system.

Mr. Williams is quite right in pointing out that ground equipment exists that is substantially better than called for in the I.C.A.O. specification, and this is most welcome and necessary in its role as an approach aid if present approach limits are to be maintained with modern aircraft. As a landing aid, however, there is the problem of over-flying aircraft to which Mr. Williams refers, and also the problem of the stability of the airborne equipment in addition to the ground equipment. A combined stability from both of something like 3 min of arc is required.

As is mentioned in the paper, the effects of earth loops, metal objects, etc. diminish with increased distance of the aircraft from the leader cables. Cars and lorries have no observable effect when they are put over the leader cable alongside the runway, and since a road must be some considerable distance from touch-down there will be no effect from this source.

This paper was also presented at meetings of Local Sections of the Institution in Glasgow and Edinburgh on 24th and 25th March, in Bristol on 26th October and in Farnborough on 22nd November, 1960.

INSTITUTION ACTIVITIES

The Committee on Broadcasting

Evidence being submitted to the Government Committee on Broadcasting has, according to press reports, been sponsored by a large number of organizations representing secular, commercial and other interests. By reason of the invitation extended to the Institution† it is understood that the Committee is also primarily concerned with the technical considerations affecting the future of radio and television broadcasting. Reports drafted by the Technical Committee and Television Group Committee of the Institution show divergence of technical opinion on the systems which might be the future basis of broadcasting, particularly in television.

In view of these differences, the Council of the Institution decided, at a meeting on 11th January, that it was not easy to meet the request of the Pilkington Committee to submit unanimous technical opinion within the time limit imposed. Apart from secular or other commercial interests, it is believed that the Government Committee will be influenced by an evaluation of the technical problems involved.

Accordingly the Council has decided, on the recommendation of both the Technical and Television Group Committee, that the Institution's submission to the Government Committee will be most helpful if it contains appraisal of several technical points of view. The Institution's standing Committees are therefore combining in preparing a report on the future of radio and television broadcasting services which will be made available to the Pilkington Committee. A general meeting of the Institution may be convened in order to discuss this report, and individual members who wish to submit *technical* argument considered are invited to submit contributions not later than 31st January 1961.

Symposium on Electronic Instrumentation for Nuclear Power Stations

The Programme and Papers Committee is organizing a half-day Symposium on "Electronic Instrumentation for Nuclear Power Stations". This will be held on 29th March at the London School of Hygiene and Tropical Medicine, at 3 p.m.

The full programme will be published shortly. At least three papers will deal with the requirements of the Central Electricity Generating Board, and describe the instrumentation for the stations at Berkeley and Bradwell in Great Britain and Latina in Italy. Additional papers will deal with other types of reactor installations. An invitation to chair the meeting has been accepted by Colonel G. W. Raby, C.B.E. (Vice-President), the Managing Director of Atomic Power Constructions Ltd.

† *J.Brit.I.R.E.*, p. 802, November 1960.

Medical and Biological Electronics Group

The Medical and Biological Electronics Group Committee has set up a working party with the Postgraduate Medical School of the University of London to organize a symposium on "Electronic Instrumentation for Cardiac Surgery". The working party comprises Mr. W. J. Perkins and Dr. C. F. Joslin from the Group Committee, and Drs. J. P. Shillingford and D. G. Melrose of the Postgraduate Medical School.

Arrangements are being made for a half-day to be devoted to the Symposium which will be held in the new lecture theatre at the Postgraduate Medical School on Monday, 27th March. This will be the first meeting of the Medical Electronics Group to be held in a hospital and organized in collaboration with the medical staff. Further details of the programme will be announced in the February *Journal*.

The Education (and Training) Committee

Publication of the Education Committee's report on "The Education and Training of the Professional Radio and Electronics Engineer"† received widespread support. Comments were considered at a meeting of the General Council on 6th December 1960. The ensuing discussion resulted in the Council deciding that the Education Committee should be re-formed as "The Education and Training Committee" working through two panels: (1) the Education Panel being primarily concerned with education policy; (2) a Training Panel being concerned with the practical training of professional engineers, technicians and craftsmen.

The work of the Training Panel is to be the subject of a further report, which it is hoped to publish in the next few months. The Panel is at present collecting evidence on present policies in industry and Government departments for the training of personnel for the various levels of employment of radio and electronic engineers and technicians. The chairman of the Training Panel is Professor D. G. Tucker (Member).

Joint Symposium on Computer Control of Air Traffic

For the first time, two of the Institution's Specialized Groups are to collaborate in holding a joint meeting. The Computer Group and the Radar and Navigational Aids Group are to hold a joint symposium in May on "Computer Control of Air Traffic". The papers will deal with the influence of man-machine relations on the input, output and display of data from an A.T.C. complex and the internal organization of a data processing equipment to fulfil the requirements.

Details of the arrangements for the meeting will be sent in due course to members of both Groups.

† *J.Brit.I.R.E.*, September 1960.

U.D.C. 621.385.832

The Cathode Loading Limit in Circular Beam Electron Devices

By

HILARY MOSS,

PH.D., (*Member*)†

Summary: On the assumption that an electron gun of rotational symmetry is operating at its theoretical Langmuir limit, the work of Haine and Schwartz is extended and combined to develop a semi-universal curve which defines its maximum usable cathode loading before the Coulomb forces operate to degrade substantially the spot. This curve shows that quite low cathode loadings (order of 0·1 amp/cm²) are sufficient to cause severe beam focus degradation, due to space charge, for the spot sizes common in normal cathode ray tubes (order 1 mm diameter). This may explain why television cathode-ray tubes are subject to spot "blooming" at higher grid drive levels.

1. Introduction

The general effects of thermal emission energies and space charge on image resolution in cathode-ray-tube-like devices have long been recognized and analysed. [1,2,3] These analyses have treated each limitation separately. No accurate treatment of their combined effects in c.r.t.'s has been published. This situation is justified partly by the extreme complexity of an exact simultaneous consideration of the two phenomena, and partly by the fortunate fact that the transition region between dominance of either effect is quite abrupt. It is easy to show that c.r.t. devices are either almost wholly thermally-limited or almost wholly space-charge-limited except under a very narrow range of operating conditions. It has not been thought worthwhile to investigate the complex situation involving both limits together since the range of such operating conditions is so small.

On the other hand it is clearly of great importance to know exactly when we reach the boundary conditions between these two sets of limit operations. M. E. Haine[4,5] seems to have been the first to outline a simple method of doing this. However his analysis is based on a very approximate treatment of the space-charge-limit equations. In particular the expression he uses for the limit current before onset of severe Coulomb repulsions (eqn. (9), ref. 5) involves a denominator which goes to zero as the parameters pass through certain entirely realistic practical values. This is a little worrying although his conclusions seem correct.

At about the time of Haine's publications, J. W. Schwartz[3] produced what must be regarded as the most elegant solution of the space charge part of the problem. Before the appearance of his work, the manipulation of the space charge equations (involving non-analytic functions) was somewhat awkward. By means of an elegant normalization procedure he deduced a universal curve permitting the space charge limit spot size to be readily calculated for a very wide range of operating conditions.

An extension of his normalized curve (Fig. 1) is given in the Appendix together with some comments on his analysis. In what follows we simply combine the analysis of Schwartz and Haine to deduce the maximum value of cathode loading which can be used before space charge repulsions at the screen begin to limit resolution.

2. Outline of Method

We follow the notation of Schwartz given in Appendix 1. Then from his normalization procedure we have

$$\frac{I^{\frac{1}{2}}}{V^{\frac{3}{4}}}\cdot\frac{z}{r_i} = \mathrm{f}\left(\frac{r_s}{r_i}\right) \equiv \mathrm{f}(p) \qquad \ldots\ldots\ldots(1)$$

where $\mathrm{f}(p)$ is the normalized function plotted in Fig. 1. For values of p above about 0·04 this function becomes doubled-valued. We are concerned only with the upper curve relating to the conditions under which the smallest possible spot is formed at the screen. In these conditions a plane of minimum cross-section (radius r_0) exists between the screen and the lens.

Assuming now a uniform spot density of value ρ_0

$$\rho_0 = \frac{I}{\pi r_s^2} \qquad \ldots\ldots\ldots(2)$$

Substituting (2) in (1) gives

$$\frac{\sqrt{\pi\rho_0}}{V^{\frac{3}{4}}}\cdot r_s\cdot\frac{z}{r_i} = \mathrm{f}(p) \qquad \ldots\ldots\ldots(3)$$

Equation (3) defines the maximum value of ρ_0, the focused spot density which can be obtained in a spot radius r_s on account of the space charge forces.

† Westinghouse Electric Corporation, Elmira, New York, U.S.A.

But by Langmuir's analysis[1], the density limit due to thermal emission velocity spread is

$$\rho_0 = \rho_c \cdot \frac{eV}{kT} \sin^2 \theta \qquad \text{.........(4)}$$

$$\left\{\frac{eV}{kT} \gg 1\right\}$$

$$\simeq \rho_c \cdot \frac{eV}{kT} \cdot \frac{r_i^2}{z^2} \qquad \text{.........(5)}$$

$$\{\sin \theta \simeq \tan \theta\}$$

Substituting (5) in (3) then gives the result sought:

$$\sqrt{\frac{\pi e}{kT}} \cdot \sqrt{\frac{\rho_c}{V^{\frac{1}{2}}}} \cdot r_s = \mathrm{f}(p) \qquad \text{.........(6)}$$

Using the normalized Schwartz curve of Fig. 1 we have plotted limiting spot size r_s as a function of cathode loading ρ_c in eqn. (6) for $V = 10$ kV and $T = 800°$C with initial beam radius r_i as a parameter. This is shown in Fig. 2.

3. Discussion

In the interests of descriptive realism we have not normalized the plot of Fig. 2 in regard to beam voltage. Normalized curves suffer from the drawback that they convey no immediate meaning. In any case we see from (6) that beam voltage enters only as a fourth root and is hence not a very significant variable.

The only approximation arises from eqn. (2), which implies a uniform spot density. In eqn. (4), ρ_0 is the peak, on-axis-density, of a Gaussian distribution. If r_s is defined as the beam width at e^{-1} amplitude then the resulting error will be small in the context of the problem. Appendix 2 justifies this step.

The surprising factor about Fig. 2 resides in the low values of cathode loading allowable before the onset of space charge limitations at the screen in the case of "normal" spot size c.r.t.s—for example of the television type. Spot sizes of the order of $r_s = 0{\cdot}4$ mm at e^{-1} amplitude would be representative. Figure 2 shows corresponding cathode loading limits at 10 kV of only the order of $0{\cdot}1$ amp/cm^2. Of course this assumes an electron gun giving 100% Langmuir limit which is not attained in practical television tubes on account of aberrations in the triode with the wide beams being used. However it seems likely that at the higher brightness and drive levels, where instantaneous loadings of several amps/cm^2 are reached, that a good deal of the commonly observed spot swelling is due to Coulomb forces at the screen.

This analysis ignores possible space charge effects at the crossover.

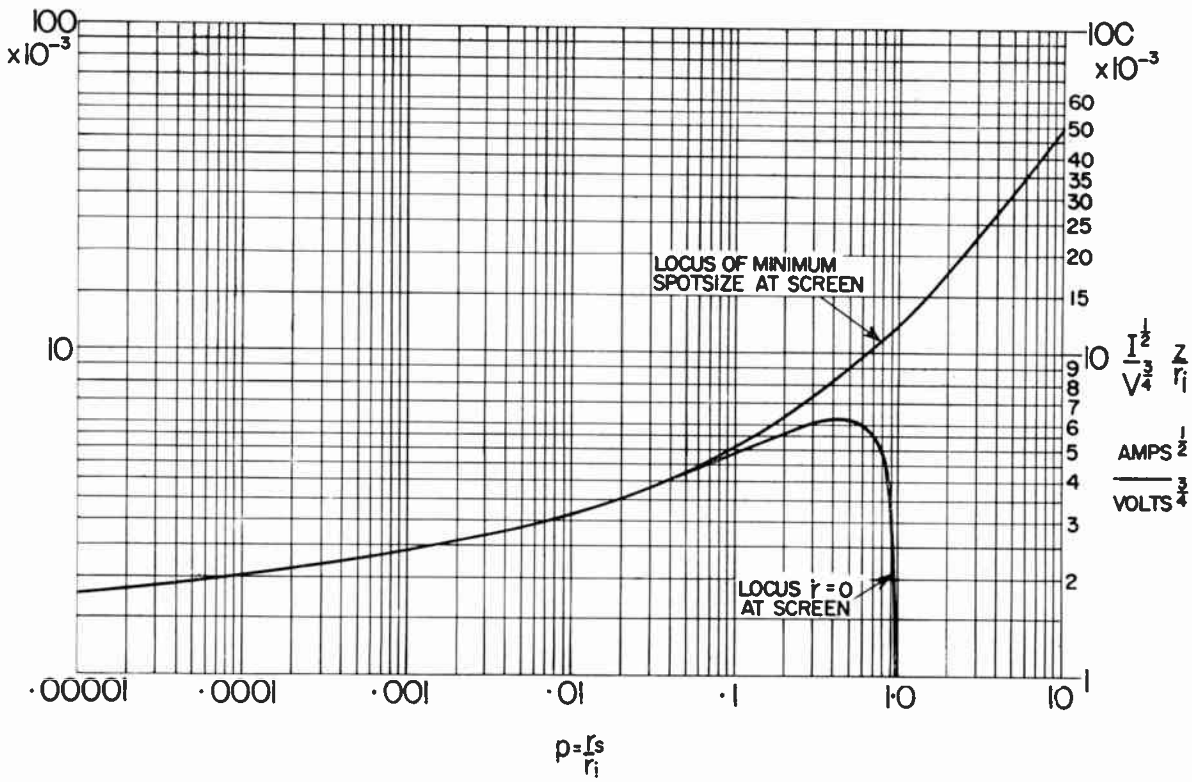


Fig. 1. Schwartz' normalized curve of $\frac{I^{\frac{1}{2}}}{V^{\frac{3}{4}}} \cdot \frac{z}{r_i}$ against p.

4. Acknowledgment

The work herein described was partly supported by United States Air Force contract AF 33(616)6219.

5. References

1. D. B. Langmuir, "Theoretical limitations of cathode-ray tubes", *Proc. Inst. Radio Engrs*, **25**, pp. 977–91, August 1937.
2. B. J. Thompson and L. B. Headrick, "Space charge limitations on the focus of electron beams", *Proc. Inst. Radio Engrs*, **28**, pp. 318–24, July 1940.
3. J. W. Schwartz, "Space-charge limitation on the focus of electron beams", *R.C.A. Review*, **18**, pp. 3–23, March 1957.
4. M. E. Haine and M. W. Jervis, "The ultimate performance of the single-trace high-speed oscillograph", *Proc. Instn Elect. Engrs*, **104B**, No. 16, pp. 379–84, July 1957 (I.E.E. Paper 2306 M. February 1957).
5. M. E. Haine, "The triode system of the cathode-ray tube electron gun", *J. Brit. I.R.E.*, **17**, pp. 211–6, April 1957.
6. H. M. Terrill and Lucile Sweeny, "An extension of Dawson's table of the integral of exp x^2", *J. Franklin Inst.*, **237**, p. 495, June 1944; and "Table of the integral of exp x^2", *Ibid.*, **238**, p. 220, September 1944.

6. Appendix 1

6.1. *Comments on the Space Charge Analysis of J. W. Schwartz*[3]

LIST OF PRINCIPAL SYMBOLS USED

(Agrees with Schwartz' notation.)

r_i initial radius of beam in anode plane

r_0 electron beam radius at point of minimum cross-section

r_s electron beam radius at screen

z distance from anode to screen

I beam current in *amperes*

V beam voltage (volts)

η electronic charge/mass ratio ($= e/m = 1{\cdot}76 \times 10^{11}$ coulomb/kilogramme)

ε permittivity of vacuum ($= 8{\cdot}85 \times 10^{-12}$ farads/metre)

$K \equiv \left\{\dfrac{\eta I}{\pi\varepsilon\sqrt{2\eta V}}\right\}^{\frac{1}{2}}$ metres/sec

$U_s \equiv \sqrt{\ln p + W^2}$

$W \equiv \dot{r}_i/K$

$p \equiv r_s/r_i$

$U(W) \equiv \int_0^W \exp x^2 \, dx$ (a tabulated non-analytic function)

k Boltzmann constant

T cathode temperature °K

ρ_c peak cathode loading in amps/cm²

ρ_0 peak density in electron spot

θ semi-angle of convergence of electron beam into focused spot

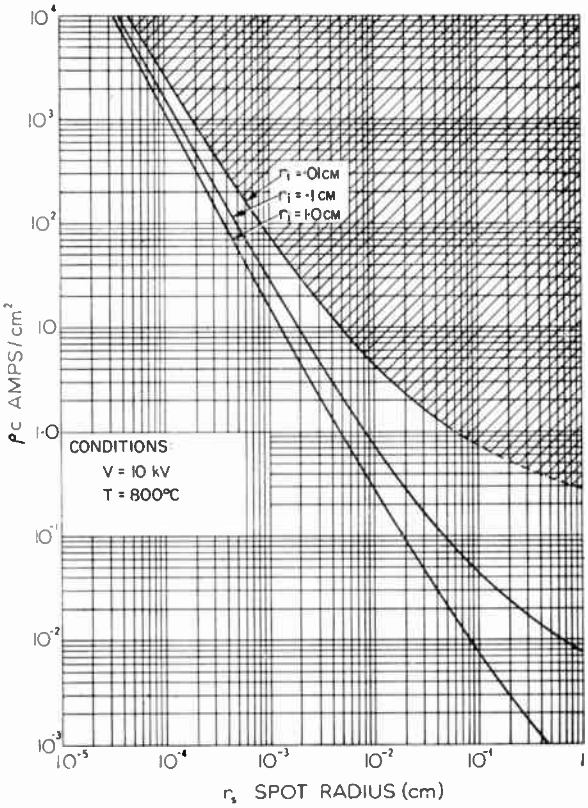


Fig. 2. Limiting spot size as a function of the cathode loading.

The problem treated by Schwartz, and his assumptions, exactly correspond to the analysis of Headrick and Thompson.[2] His paper differs only in an ingenious normalization procedure which presents the results in a most convenient form.

Schwartz in his eqn. (24) gives the condition for minimization of r_s as

$$2W.\exp - W^2[U(W) + U\sqrt{\ln p + W^2}] = 1 + \frac{pW}{\sqrt{\ln p + W^2}} \quad \text{.........(24)†}$$

This defines the W–p relationship needed to minimize r_s. He does not plot the solution of this transcendental equation, nor does he discuss it, beyond remarking that it has to be treated graphically.

Some analysis of eqn. (24) might however be possible if we could approximate the non-analytic function $U(W)$ in closed form. Over a limited range of W this can be done since for $W > 1{\cdot}5$

$$U(W) \equiv \int_0^W \exp x^2 \, dx \simeq \frac{\exp W^2}{2W}$$

† A root sign over the term $\ln p + W^2$ was omitted by printer's error on the left-hand side of his equation.

so that we may write:

$$U(W) = \frac{\exp W^2}{2W} + \delta \qquad \text{.........(7)}$$

Now suppose that

$$\ln p = -W^2 + \lambda \qquad \text{.........(8)}$$

is a solution of eqn. (24). If this is to be true then by substitution of (8) and (7) in (24) we must have

$$2W.\exp - W^2 \left[\left(\frac{\exp W^2}{2W} + \delta\right) + U(\sqrt{\lambda})\right] = 1 + \frac{W.\exp(-W^2).e^{\lambda}}{\sqrt{\lambda}}$$

i.e. $$1 + 2\delta W \exp(-W^2) + U(\sqrt{\lambda})2W \exp(-W^2) = 1 + \frac{W.\exp(-W^2).e^{\lambda}}{\sqrt{\lambda}}$$

i.e. $$2\delta + 2U(\sqrt{\lambda}) = \frac{e^{\lambda}}{\sqrt{\lambda}} \qquad \text{.........(9)}$$

At first sight this condition looks almost as intractable as the original eqn. (24), but this is not so. The function $U(W)$ has been tabulated.[6] Some values are given in Table 1 which also shows the corresponding approximation $(\exp W^2)/2W$ and the difference δ.

Table 1

W	$U(W)$	$\frac{\exp W^2}{2W}$	δ	λ (Smaller root)
1	1·46	2·15	−·69	no real value
1·5	4·06	3·16	+·90	no real value
2·0	16·45	13·65	+2·8	·034
2·5	115·6	103·6	+12	·001 74
3·0	1444	1350	+94	·000 028 3

It can be seen that δ rises rapidly for $W > 2$. Now $e^{\lambda}/\sqrt{\lambda} \to \infty$ as $\lambda \to 0$, so that if δ is large, *one* solution of (9) will always be found for a value $\lambda \ll 1$. In these conditions since $U(W) \to 0$ for $W \to 0$, we may rewrite the condition (9) in the simpler form

$$2\delta = \frac{e^{\lambda}}{\sqrt{\lambda}} \qquad \text{.........(10)}$$

Again for $\lambda \ll 1$, $e^{\lambda} \to 1$ and we may solve (10) for λ giving

$$\lambda \simeq \frac{1}{4\delta^2} \quad (\lambda \ll 1) \qquad \text{.........(11)}$$

Table 1 includes values of λ and shows that for $W \geqslant 2$, λ is very small. Thus we find that a partial-range solution of eqn. (24) is simply

$$\ln p = -W^2 \qquad \{W \geqslant 2\} \qquad \text{.........(12)}$$

This is indicated in the plot of Fig. 3. The remainder of the solution of (24) for $0{\cdot}4 \leqslant W \leqslant 2$ is also given in Fig. 3. This was obtained by an iterative numerical analysis.

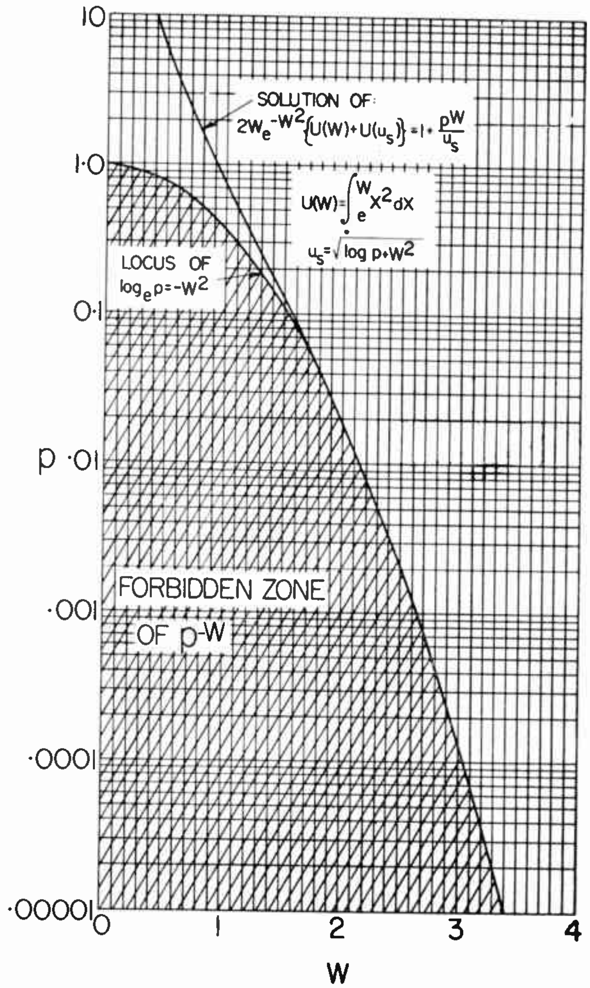


Fig. 3. Partial range solution of Schwartz' equation for the minimization of the electron beam radius at the screen.

6.2. *Determination of the Single Branch Portion of the Normalized Space Charge Curve* (Fig. 1)

For values of $W > 2$ where the simple minimization of eqn. (12) applies, the calculation of the normalized curve of Fig. 1 is very simple.

Starting from Schwartz' eqn. (27)

$$\frac{Kz}{\sqrt{2\eta V}} = 2r_i \exp -\beta^2[U(\beta)+U(u_s(\beta))]$$

we substitute from (12) so that $U(u_s(\beta)) = 0$.

Here β is the value of W satisfying the normalization condition (12). Substituting again from (12) then gives

$$\begin{aligned}\frac{Kz}{\sqrt{2\eta V}} &= 2r_i \exp -W^2[U(W)] \\ &= 2r_0 . U(W) \\ &= 2r_s . U(W) \qquad \text{.........(13)}\end{aligned}$$

In this case $r_s = r_0$ at the screen surface. Given p we find W from Fig. 3. $U(W)$ is then found from its tabulation. Equation (13) then permits calculation of z for any assumed r_i value. Thus the normalized ordinate $\frac{I^{\frac{1}{2}}}{V^{\frac{3}{4}}} \cdot \frac{z}{r_i}$ may be calculated.

7. Appendix 2

Here we are simply concerned to show that the radius of a right cylinder having the same height and volume as the solid Gaussian distribution it envelops, is given by the radius of the Gaussian solid at its e^{-1} ordinate.

Let the Gaussian solid be defined as the distribution:

$$\rho = \rho_0 \exp -Br^2 \qquad \text{.........(14)}$$

Let the radius of the enveloping cylinder of equal volume be R. Then the volume of this cylinder is

$$V = \pi R^2 \rho_0 \qquad \text{.........(15)}$$

Now the volume of the whole Gaussian solid is

$$2\pi\rho_0 \int_0^\infty r . \exp -Br^2 = \frac{\pi}{B} . \rho_0 \qquad \text{.........(16)}$$

Equating (15) and (16) gives

$$R^2 = \frac{1}{B} \qquad \text{.........(17)}$$

Substituting the condition (17) in (14) then shows that R corresponds to the condition $\rho/\rho_0 = e^{-1}$ as we were required to show.

Manuscript received 28th April 1960 (Paper No. 600).

News from the Commonwealth Sections

Increased activity is reported from all the Institution's Sections in the Commonwealth. In the February and March Journals, *accounts will be given of membership activities in Australia, New Zealand, Pakistan and South Africa. It is understood that a number of members from the Commonwealth Sections will be attending the 1961 Convention, details of which are given on pages 93–4 of this* Journal.

CANADA

The Montreal Section Committee has arranged a most interesting variety of meetings for the current session. The programme started in October when Mr. D. F. Gilvary (Graduate) spoke on "The Mid-Canada Early Warning System", laying particular emphasis on the semi-technical aspects of detection and communication with co-ordinated areas. Ancillary services necessary for the maintenance of the system were also reviewed. He concluded with an account of the life of the engineer in Northern Canada where climatic conditions present problems in engineering and sometimes of existence itself.

Members of the Section visited the TCA "Vanguard" flight simulator installed at the new TCA Maintenance Building, Dorval, on 10th November. The visit proved to be most worthwhile and interesting as it gave members an opportunity for discussions with the engineers responsible for the design and installation of this simulator.

On 23rd November Mr. D. W. Grierson's paper on "Printed Circuit Techniques" dealt in detail with methods of producing printed circuits both in prototype quantities and on a production basis. Subjects covered included theory, economics, quality, utilization, packaging and the advantages of printed circuits. His paper concluded by reviewing less familiar applications and indicated future possibilities for the printed circuit.

The Secretary to visit Canada.—All members in Canada will be pleased to learn that the Council has agreed to the request of members that a further visit to the Canadian Sections should be made by an officer of the Institution. It is felt that the Section in Montreal might well operate independently, whilst still maintaining an overall consultative Canadian body on the lines of the present Advisory Committee. This and other matters were discussed at a meeting of the General Council in London on 11th January 1961, when it was resolved that the General Secretary, Mr. Graham D. Clifford, should visit Canada in the early autumn of 1961.

Council has also received a request from members and industrial organizations in the Eastern States of the U.S.A. to have an opportunity of discussing Institution affairs. It is hoped that Mr. Clifford will be able to visit New York after his visit to Toronto, Montreal and Quebec.

INDIA

In recent months there has been an increase in the number of papers submitted by members in India. Several have been approved for publication by the Papers Committee and one of them is published in this *Journal* (pp. 49–56). The Papers Committee emphasizes that whilst all papers are considered for the annual award of the twenty premiums and awards now given by the Institution, only papers from Indian nationals are considered for the award of the Sir Jagadis Chandra Bose Premium. Because of limitations on *Journal* space in 1960, some papers which were submitted by Indian members could not be considered for the award of the 1960 Bose Premium. The Council is therefore considering a proposal that two Bose Premiums be awarded this year.

Administration.—Much of the correspondence received from Indian members and enquirers, particularly applications for regulations and forms for election, transfer and examination entry, is now handled by the office of the Indian Advisory Committee, P.O. Box 109, Bangalore 1. The office is under the supervision of the Honorary Secretary, Colonel B. M. Chakravarti (Member). For some years now, the regulations of the Institution governing membership and examination and the application and proposal forms for India have been printed in Bangalore.

The Chairmen of the five Indian Sections—New Delhi, Bombay, Calcutta, Bangalore and Madras—comprise the Indian Advisory Committee, under the Chairmanship of Major General Brahm D. Kapur (Member). The Advisory Committee is to meet within the next few months and, with the approval of the General Council, will make recommendations on the possibility of holding a Convention of the Institution in New Delhi towards the end of 1962.

Government of India Recognition.—Additional to the information published in the History of the Institution† is the announcement of January 1961 that the Government of India has now accorded recognition to membership of the Brit.I.R.E.—see page 4 of this *Journal*.

The Institution's representatives in these discussions with various Indian Government Departments have been Major General B. D. Kapur and Col. B. M. Chakravarti.

† "A Twentieth Century Professional Institution"—The Story of the Brit.I.R.E. Published by the Institution, price 30s. (Rs. 20), post free.

U.D.C. 621-5

System Engineering in Theory and Practice

Presented at the South Western Section's Convention on "Aviation Electronics and its Industrial Applications", held in Bristol on the 7th–8th October 1960.

By

M. JAMES,
DIP. EL. (HONS.), (*Member*)†

AND

G. S. EVANS, (*Associate Member*)†

Summary: An account is given of the many areas of science and engineering that are involved in the complex fields of aircraft and missile engineering. The manner in which system engineering allows co-ordination of these activities into a coherent scheme is outlined, together with an indication of the theoretical tools at the disposal of the System Engineer.

A detailed description is then given of an "on-line" process control analogue computer of novel principles. An arithmetic unit consisting of a minimum number of computing amplifiers and multipliers is used to solve the equations in a piece-wise fashion according to a pre-set programme in which each step is selected sequentially by a switch system. Stores are used to retain information on the result of a step in the solution, these data being used in a later stage of the computation.

Some trends for the future are briefly examined and a forecast is made of certain subjects which will merit further study, if the rapid progress of automatic control systems is to be maintained.

1. Introduction

In a large and complex industry such as Aircraft and Missile Engineering many specialist areas of endeavour are to be found. There will certainly be found Mathematical, Aerodynamic and Thermodynamic Services, Radio, Electronic and Electrical Groups, Physics, Chemical and Optical Laboratories, Precision and Heavy Mechanical Engineering, Hydraulic and Pneumatic Actuator Groups, Automatic Control Laboratories and Computing facilities using analogue or digital machines. These may be combined under single headings or be even further sub-divided. A further complication may arise if a Company is organized on a project basis for then one finds these teams repeated in each of the projects with more or less emphasis upon individual activities depending upon the project.

These separately definable activities will be, in general, under the control of a specialist Group Leader, a man well versed in his particular subject. His Group may be working to specifications which will demand a particular performance of his specialized device.

The specifications will originate from a System Engineer who is responsible for combining several devices into a single working system. For instance, a guidance system may consist of a radar transmitter/receiver from the Radar Group with the aerial mounted on a gyro from a Precision Instrument Group. Signal amplifiers may be supplied by an Electronic Group and power from an Electrical Group. Here then is the first instance of a System Engineer—he is a man able to understand in some detail a number of engineering sciences, but in particular able to see how one device can work in conjunction with another.

† De Havilland Propellers Ltd., Hatfield, Hertfordshire.

2. A Complex System

This theme may be carried a stage further by considering a complex of interconnected systems. A good example of this is a ballistic missile. Here one may find guidance and control systems, power, telemetry and safety systems, propulsion and pressurization systems, and airframe systems, together with a warhead system. The systems can be considered separately each having discrete inputs and outputs, in fact the classical "black box". These systems pose the usual problems of stability, performance, reliability, size, weight, etc. But now consider the greater complication, all these complex systems must work in harmony together to form a weapon system: again the considerations of stability, reliability, performance, etc. But the problem is not finished, a ballistic missile has a ground system of equal or greater complexity and the System Engineer has to design the final amalgam of weapon and ground system.

One now has a vast feedback control system making up the ballistic missile which ranges from heavy civil engineering to the smallest transistor circuit. It has linear and non-linear elements. A mass of data must be collected and automatically checked. Launching involves precision measurement at high speed, sequencing of events according to a programme sometimes incorporating permissive progress on the basis of several variables. To complicate matters human decision elements probably appear in the loop.

3. The System Engineer

To bring order to this scene is the role of the System Engineer. He might be classed as a scientific and engineering general deploying many areas of knowledge. To help him, he has various mathematical disciplines and computational facilities.

The prime consideration of system engineering must be integration, all elements of the system must combine together for some common purpose. In fact all components in the system must contribute to the production of a set of optimum outputs as a result of a given set of inputs.

Most large-scale systems involve interaction either by unavoidable cross coupling loops or as a natural result of the system dynamics wherein a change in one variable has a first-order effect on another.

The concept of feedback is fundamental to very many systems. Regarded as a servo, a system may appear most complex involving many closed loops with subsidiary loops. From the point of view of servo theory further complications may arise when parts of the system appear incapable of description by simple functions, but may need statistical methods to describe them. In some systems a computer, either digital or analogue, may appear as a vital component. This factor after all is the essence of true automatio n as opposed to mechanization, and a typical on-line machine will be considered later.

The function of the human must be recognized—in some systems he might be used as a transfer element—a reliable servo with limitations, in others as a decision-making element—a slow-speed computer. In particular he must be fitted to a system in the optimum manner; this study leading one to the relatively new idea of human engineering or ergonomics.

4. System Design

System design can be isolated into two areas—the exterior and the interior. By exterior system design is meant the statement of the problem, i.e. the requirement on the system taking account of its environment. Interior system design is concerned with the system, i.e. the synthesis of a solution including equipment, procedures, and people.

4.1. *Exterior Design*

Considering exterior design in more detail—the first task is to obtain a statement of the problem. This obvious stage is not always easily fulfilled. It is perhaps relatively easy to specify the performance required of a data processing system, form of input signal, measuring accuracy, speed, outputs, and so on, but to formulate the problem involved in the control of a steel-mill, or the automatic scheduling of a machine shop has far more intangibles. It is important that the system engineer should completely understand the problem in its fundamental terms, because in taking the broad view he may observe a mis-statement of the problem if it was set from a narrow viewpoint. This can often happen if a problem is set by one department of a large organization, since here the solution often involves the statement of a much broader problem embracing several other departments or parts of the process. One of the tools of problem statement is operational research; this becomes of particular use when considering complex structures such as a factory administration system where transfer of information and materials takes place over many interacting channels.

Having obtained a full grasp of the problem statement, it is sometimes possible to produce a mathematical model describing it. The School of Industrial Dynamics, a part of M.I.T. at Boston, Massachusetts, have recently produced two large-scale models of this type. In the first instance they simulated the American Boot and Shoe Industry in all its aspects of economics, standards, materials, costs, labour relations, and—most surprisingly—customer reaction to new fashions. The second example was even more complex since they claim to have set up a digital representation of the American Economy. These models are of no use without data and the third phase of exterior design is design of experiments to obtain this data. In fact the model cannot in many cases be commenced without data and one is faced with a closed loop in this part of the design.

To help in the design and analysis of the experiments, mathematical statistics can be called into play. When obtaining the data a better impression can be gained of the accuracy to be expected on the inputs to the system and this may modify the original specification. For instance on a chemical plant there is little point in using a digital computer having an accuracy of $1:10^4$ as part of the control system, if the input information can only be measured to 1% or 2%. Subsequent to the design of the experiment is its actual performance in the field. These experiments may involve complex equipment such as high-speed wind tunnels, pilot plant, and data loggers. Data loggers are particularly useful in conjunction with mathematical statistics to help in analysis of complex and non-linear processes. On many chemical plants frequency or transient analysis prove impossible because of extremely long time constants, large attenuations, or random disturbances.

Under these circumstances information may be obtained with a digital data logger recording on magnetic tape and subsequently being reduced, reconstituted, linearized and correlated by means of a data processing computer.

4.2. *Interior Design*

The interior system design may be broadly classified under one of three headings: single thread design, high traffic design and competitive design.

In a single thread system there is only one route for each piece of input information; it is possible to describe every response that they will produce in the system.

This is equivalent to saying in servo parlance that all transfer functions are known and are unique. Therefore it is always possible to describe the outputs of the system in terms of its inputs. Examples of this type of design are commonplace. The simpler guided weapons, the automatic machine tool, the conventional process control systems are all single thread designs. System logic is the tool used to aid design and results in the well-known block diagram which determines the individual system components and the rules for their connection.

High traffic design is concerned with systems having many inputs where these inputs are distributed in time in the probability sense. If one has a system of fast response compared to the statistically-probable interval between inputs then no delay of information handling will occur. If, however, it is possible for an input to arrive while the system is still responding to the previous input then a queue will form. In the limit this queue would become of infinite length and the system would be overloaded. The choices are then more channels of service (tending to single thread design), faster response of a single channel, or buffer storage to accommodate queues at peak periods. The proper choice is the essence of high traffic design, one of the tools being Queuing Theory. To simplify the application of this theory it is usually assumed that inputs are either equi-spaced in time or follow a Poisson distribution, and that the holding time or channel time constants are constant or exponentially distributed. Given these assumptions then a digital computer may be used to simulate the queuing problem using a technique known as the Monte Carlo method. High traffic systems are fairly commonplace, examples being, of course, telephone exchanges, aerodromes, transport systems and some forms of alarm recorders used on process plants. This design method warrants careful attention for it can lead to the optimum use of expensive equipment in many applications where an intermittent or sampled data output can be tolerated.

Competitive design involves systems where a maxima or minima is being continuously sought. A branch of mathematics known as the theory of games can be used and has been applied particularly to military systems. A so-called "war game" has been set up on a digital computer in the United States and has investigated many forms of attack and defence situation. A simplification of this method can be used in the design of optimizing systems for use in industry. Here the process is continuously investigated by an on-line computer to decide the optimum settings for a range of variables to achieve the most favourable quality/cost ratio on the finished product.

4.3. *Other Theoretical Tools*

Other disciplines which find their place as tools at the disposal of the System Engineer are:

(*a*) Servo Theory, both linear and non-linear.

(*b*) Information Theory to help with communication problems.

(*c*) Linear Programming. Akin to game theory and helpful in the analysis of complex problems of many variables.

(*d*) Group Dynamics to assist in the study of information flow in human organizations.

(*e*) Cybernetics to marry together the fields of biology, physiology and engineering.

5. Comparison of the Aircraft Industry and the Process Industries

In applying automatic control and system engineering experience as understood by the aircraft industry to the problems of process plant, several points become apparent.

5.1. *System Dynamics*

The levels of understanding of system dynamics are vastly different. The System Engineer in the aircraft industry can generally express the transfer functions of all his components down to the last nut and bolt.

The System Engineer in the process industry has not in the past, however, been required to obtain the same level of understanding. It is only now with the arrival of more sophisticated control equipment, in particular computers, with demands placed upon him to control processes of greater complexity, and with an ever increasing need to obtain closer and more optimum control, that he is beginning to obtain theoretical understanding and experimental confirmation of the process dynamics.

5.2. *Linear and Non-linear Systems*

As far as possible the aircraft industry uses linear devices, endeavours to force non-linear devices to be linear even at a cost in performance, or approximates them as linear for small signal amplitudes. This is done to allow a full theoretical calculation to be made on expected performance acting as a design aid. It also allows simulators to be used to "fly" missiles on flight plans extrapolated from one or two cases proved from actual firings. On the other hand the process industry is faced with situations on plants that are undoubtedly extremely non-linear and where little if anything can be done to linearize. In these situations describing-function techniques must be used where the output response is determined by the form and amplitude of the input signal. Without a general method of analysis for non-linear systems the theoretical evaluation of these systems is handicapped and simulation made extremely difficult.

5.3. *System Time Constants*

In the aircraft industry systems have time constants from seconds to milliseconds. Many components have natural frequencies that are both high and very underdamped. These components having second-order or higher transfer functions can be difficult to stabilize when included in a control system which must also have a fast speed of response.

This state of affairs is unusual in process work where simple first-order lags predominate due to hold-up times in the plant leading to time constants between several seconds and hours. As a corollary to this, because of the dynamic state of systems in aircraft, automatic control of some sophistication must be used. On process work, however, changes are relatively slow and are within the comprehension of human operation. Automatic control must justify itself on other scores, such as better quality of product at reduced cost.

5.4. *Cross Coupling*

Aircraft and guided weapon systems usually have several inputs to the control loops with cross coupling terms both deliberate and unavoidable. Most of the inputs have first-order effects on the outputs and control is established on the basis of many variables. Compared to this one finds in the process industry that due to the predominating lags it is possible to alter dynamically (as opposed to the steady state) one variable without effecting others. Simple controllers operating on individual parameters such as flow, temperature, pressure, etc., can therefore be used without cross coupling, and attempts at multivariable control are in their infancy.

6. The Application of Computers

Taking note of these observations, and coupled with the fact that a process control system may be delineated into measurement, communication, computation and control, it was decided by the authors' company that any advancement in process control would centre on computation. The use of on-line process control computers will bring about multivariable control. They will allow quality of product to be inferred by calculation from several variables. They will be used as decision making elements for logical sequencing of plant, typically the start up of boiler and turbines. They can give operating efficiency by continuous computation of mass and thermal balances. They will be applied to the control of complex parts of plants such as distillation columns and evaporators. They will be used more and more for the automatic optimization of processes against a cost/quality criteria.

7. Justification of an Analogue Machine

Taking account of the present form and accuracy of plant instrumentation, it was felt that a start should be made with an analogue computer, although the future may well lie with digital differential analysers or digital computers or hybrid arrangements comprising both analogue and digital.

Examining the present form of analogue computing led to several conclusions. Firstly, being parallel operation machines meant that their speed of problem solution was extremely fast compared to plant time constants. Secondly, because of their scientific usage, the form of mechanical construction, setting up arrangements and general stability and reliability did not match up to demands of a process environment. Lastly, the price was too great when they had to be justified to a shrewd and often incredulous management on the basis of a few per cent saving demonstrated only on paper.

8. An On-Line Process Control Analogue Computer

From these considerations came the synthesis which led to the on-line process control analogue computer, termed "Anatrol".†

8.1. *General*

"Anatrol" uses a time-sharing principle in which the minimum number of computing amplifiers and multipliers are formed into an arithmetic unit. This unit is then used to solve the process equations in a step-by-step fashion according to a pre-set programme, in which each step is selected sequentially by a switch system. Intermediate information as a result of a step in the solution is transferred to capacitor short stores to be used at a later stage of the computation. Long stores are used to clamp the outputs at the last computed figure while new values are being determined.

† Patent applied for.

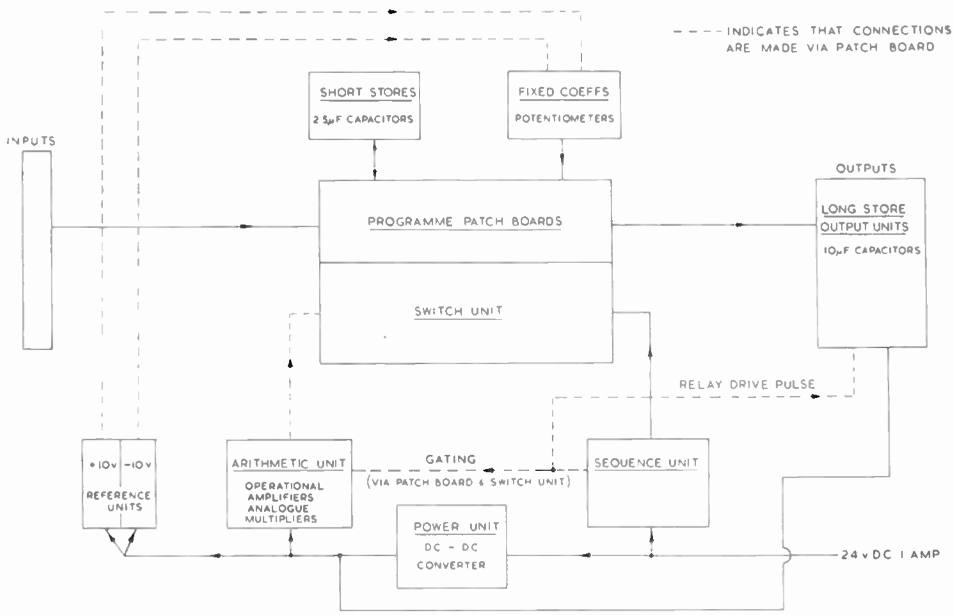


Fig. 1. General layout of the "Anatrol" on-line computer.

Figure 1 shows the general arrangement of the computer. Up to twenty-five steps are available, and with a full complement of operational amplifiers and multipliers, the computer has the potential of a conventional analogue machine employing at least seventy-five amplifiers and twenty-five multipliers.

Module construction is used in "Anatrol" to enable different combinations of circuit block to be used according to the problem, and these modules, which include the operational amplifiers and analogue multipliers, co-efficient potentiometers, voltage references, switching units, short stores, inputs and outputs, have their active connections brought out to the front via taper pin sockets. This enables a problem to be patch-connected via the 25-way switch so that each step of the switch contributes in succession to the solution of the problem.

In operation, the patching facility is sealed inside the computer to protect it from environmental conditions and accidental alteration.

8.2. *The Operational Amplifier*

Figure 2 shows the configuration of the amplifier used which was designed to a specification dictated by the unique requirements of the "Anatrol" system. In drawing up this specification it was considered of more value to lay down the amplifier requirements as a functional element of the computer than to identify it in isolation by parameters which, to mean anything, had anyway to be qualified by circuit conditions.

Thus the specification which covered an amplifier for use either as part of the arithmetic unit or an output unit was as follows:

(*a*) *Output Requirements*
± 10 V maximum, ± 30 mA maximum.

(*b*) *Temperature Environment*
0 to 40° C giving 0·1 % accuracy.
-10 to $+40$° C giving lesser accuracy.

(*c*) *Holding Integrator Error*
With a 10 μF capacitor connected between output and input terminals, the output voltage shall not change by more than 0·3 mV/sec.

(*d*) *Voltage Amplifier Error*
As an amplifier with a 1 megohm feedback and 100 kilohms summing resistor, the contribution of the amplifier to the output error shall not exceed 5 mV.

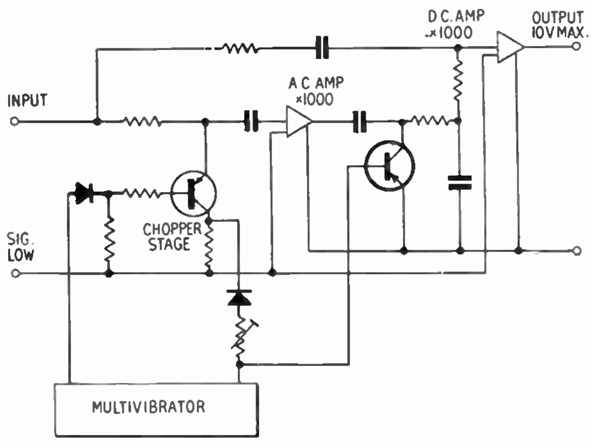


Fig. 2. The operational amplifier.

(*e*) *Speed of Response*

The amplifier output shall be capable of meeting the required accuracies within 1/100 sec.

To satisfy the specification a transfer resistance of 3×10^{11} V/A was achieved using a chopper-stabilized amplifier as outlined in Fig. 2 and an input current of 5×10^{-10}A easily permitted its use in a long store/output unit. With a closed loop gain of 10, a bandwidth of 300 c/s was accepted since,

$$f_{(0\cdot1\%)} \simeq 7 f_{(3\mathrm{dB})} \simeq \frac{7}{2\pi CR} = 117 \text{ c/s}$$

Such margins of safety were considered desirable to ease production tolerance problems.

8.3. *The Multiplier*

A varying mark-space ratio principle was chosen as a basis of design for the multiplier in the arithmetic unit since there already existed a background of experience of this method and initial investigations led to the feeling that the accuracy would be most easily met in this way. Its rather limited speed of response consistent with a small output ripple was quite acceptable in the computer.

The Hall effect and the Wilby multipliers were considered, but the former suffered at that time from large variations due to temperature, while the latter was not directly capable of accepting two d.c. signals.

An accuracy of 0·2% of full scale from half scale to full scale was laid down as a specification, and an input of ± 10 V maximum on Y and $+10$ V maximum on X was required to give $\dfrac{\pm XY}{10}$ at the output terminals. The general circuit configuration which met this specification is given in Fig. 3.

The operation of the system is such that the X input voltage V_X and the output from a bistable circuit, switching between zero and a reference voltage V_R are summed at the input of an analogue integrator. The voltage time integral at the output is then applied to a threshold-gate to initiate the bistable. Phasing is such that the system is regenerative giving a

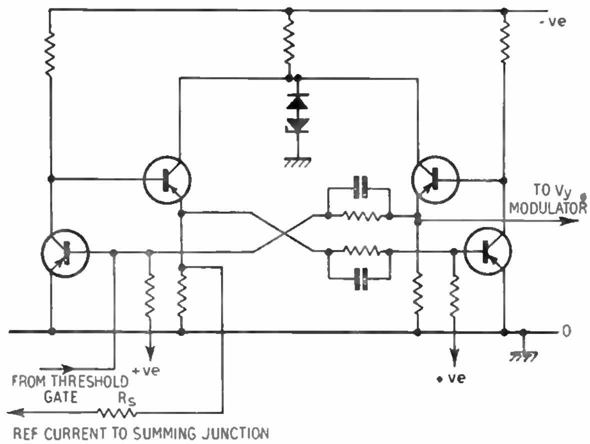


Fig. 4. Pulse amplitude stabilized bistable circuit.

pulse width modulated output from the bistable. Hence $\dfrac{t}{t+\tau} \propto V_X$, where t and τ are the dwell times of the two stable states of the bistable. This pulse train is then used to modulate the Y input voltage via a transistor switch, such that an output equal to

$$\frac{t}{t+\tau} \times V_Y = \frac{K V_X V_Y}{10} \text{ results.}$$

The constant K describes the relationship between V_X and $\dfrac{t}{t+\tau}$ which, of course, depends upon the ratio V_X/V_R maximum and the ratio R_X/R_S (the summing resistors to the integrator).

The required stability of the pulse amplitude from the bistable circuit was achieved by switching between zero and a voltage level defined by a Zener diode having a positive temperature coefficient, a forward conducting diode with its negative temperature coefficient, and the heavily bottomed transistor in the bistable. This is shown in Fig. 4.

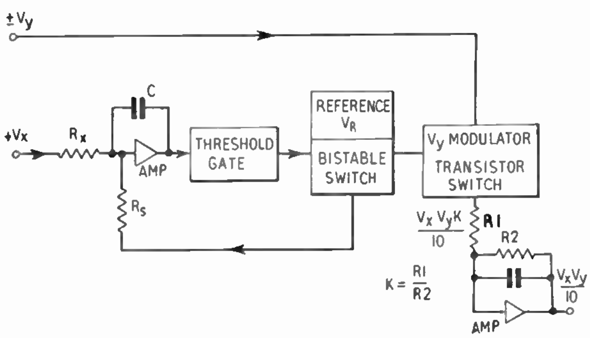


Fig. 3. The multiplier.

8.4. *The Output Unit*

In serving as a non-destructive read-out analogue store, the output unit was required to accept the result of a computation from the arithmetic unit and offer this result as a low impedance source at the output terminals of the computer, until a recomputed result was available. The computer outputs lie in the range ± 10 V and remain steady to 0·1% for holding periods of 25 sec whilst supplying load currents of up to 30 mA.

Throughout the computer it was necessary to guard against errors induced during the switching periods due to the presence of large current drive pulses to the uniselectors, and also variations in switching times of the uniselector banks—obviously a significant factor in this application. For this reason isolating relay contacts and transistor switches are used to assist accurate transfer of information to the long and short stores. Figure 5 shows the circuit arrangement for the short store case, where contact A1 is timed to open before movement of the uniselector (see sequence unit).

Similarly the output unit is equipped with a transistor switch to serve the same purpose, and the circuit of Fig. 6 shows this.

During the read-in period the relay contact B1 is closed and the transistor turned on by the negative relay drive. This enables C to be charged rapidly via the output impedance of the amplifier and the limiting resistor R. The hold condition is then obtained by removing the drive signal from VT1 and RLB/1. Since VT1 switches off before the relay de-energizes, the store is still tending to fill (although with a much longer time constant) when B1 opens, after which the uniselector is allowed to move without disturbing the circuit conditions.

A 10 μF capacitor for the long store is made up by casting twenty 0·5 μF polystyrene capacitors in polyurethane resin.

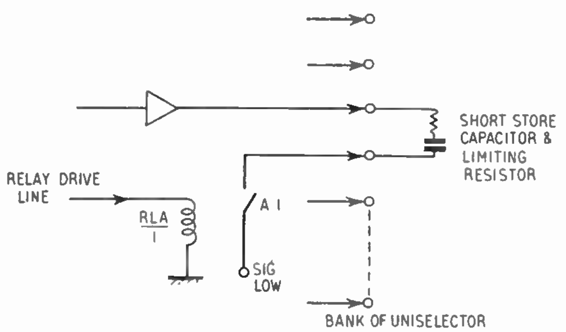


Fig. 5. Short store output unit.

8.5. *The Sequence Unit*

Two series of timing pulses are generated within the sequence unit, pulses of approximately 100 millisec duration to progress the uniselector at one step per second, and pulses of approximately 200 millisec duration overlapping the 100 millisec pulses to serve as relay drive. The timing arrangement consistent with the considerations of the previous section is shown in Fig. 7.

8.6. *The Reference Unit*

Two independent reference outputs +10 V and −10 V accurate and stable to 0·01% are derived from two independent SZT2 reference junctions. The voltages from these junctions are then matched into the 30 mA load via two stabilizer circuits, one built with *pnp* and the other with *npn* transistors, and the

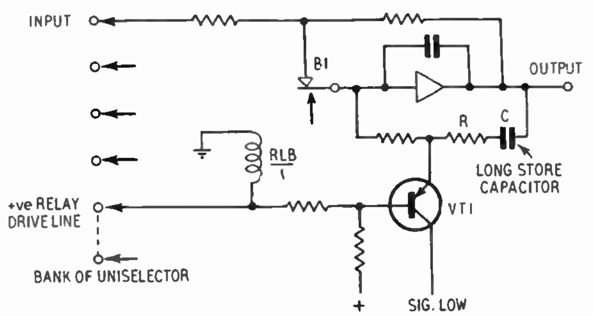


Fig. 6. The long store output unit employing a transistor switch.

temperature coefficient of the reference junctions is optimized by adjusting the current through them.

The output voltages from the reference unit are designed to be accurate and stable to 1 mV over the temperature range 0–40° C for 1% input variations and loads varying between 27 and 33 mA.

8.7. *Power Supplies*

It was decided at the outset to design the system for a 24 V d.c. power input, since this would allow emergency operation from standby batteries in the event of mains failure; a four-transistor bridge-type a.c./d.c. converter was therefore used to supply all the units excluding the sequence unit. Following the converter, series stabilizer circuits provide low impedance outputs of ±20 V and ±11 V accurate to ±1% for supply variations of (−20% to +10%) and load variations imposed by the widely different possible complements of modules.

To operate the computer from the mains a further power unit is necessary to supply 24 V at 1 amp.

8.8. *Reliability*

Much attention has been given to reliability, since it is envisaged that the machine will be operated continuously under difficult environmental conditions. This computer has therefore been shock-mounted in a strong casting provided with sealed doors which may be locked. All cable entries are hermetically sealed.

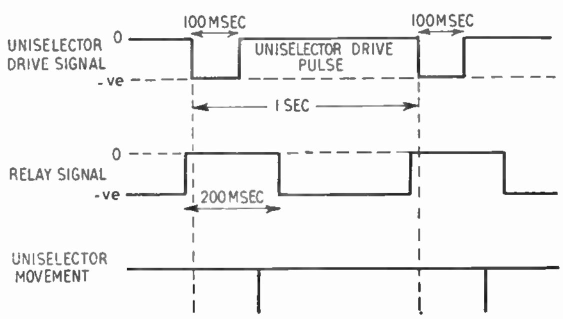


Fig. 7. The sequence unit waveforms.

The electronic units embody conservatively rated semi-conductor techniques, with components mounted on plug-in printed circuit boards (Fig. 8).

The electro-mechanical switching elements are of proved reliability. High grade, long life, low resistance contact materials are employed, and suitable choice of circuit impedance has minimized the effect of contact resistance.

As an over-riding safeguard the machine is organized to carry out a self-check on all its critical components once per computer cycle of 25 seconds. In the event of a fault being shown to exist, the computer can give an alarm and, if need be, disconnect itself from the process being controlled.

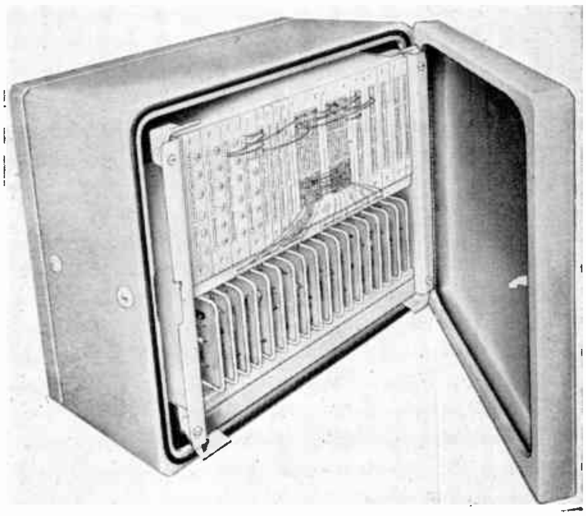


Fig. 8. The "Anatrol" process control computer.

9. Pattern for the Future

What will be the pattern of system engineering in the future? Undoubtedly the number of applications will increase at a high rate, and this will be particularly true in the realm of automatic control of plants and processes. In these areas the surface has barely been scratched and once the system dynamics have been determined there will be a rapid growth in the number and complexity of control systems.

Complexity is sure to increase, it always follows in the wake of human progress. Demands for better efficiency and closer control will become louder. Computers will replace humans more and more as decision-making elements.

9.1. *Advances in Theory*

Advances in theory and in the understanding of theory must come to the aid of the System Engineer. The disciplines of system engineering already described must be applied with greater vigour if complexity is not to overwhelm the system design study. The meaning and potential of theory must be made known to more engineers if they are to derive any benefit from the concepts of these theories.

An example is information theory where mathematical reasoning has produced extremely valuable relations between signals, noise, information content, bandwidth and entropy. Despite the many apparent applications of the theory little has been achieved in terms of practical devices. Is this because the mathematician who understands the theory cannot put it into practice, while the practising engineer makes no attempt to understand the possibilities portrayed in the theory?

9.2. *Advances in Computer Applications*

Undoubtedly computers, both analogue and digital, will have an ever more important role to play. To integrate them into his control loops, the System Engineer must study and develop the concepts of digital and sampled data servos which hold the key to higher accuracy and greater efficiency. They will require a fresh approach to the design of control actuators if the present redundancy of digital-to-analogue converters is to be avoided. In the same light, instruments must be extended to the point where cheap accurate digital transducers are available. At the very least information must be in the time-domain if accuracy is to be preserved over the long transmission paths that will form a part of the centralized controls of the future.

9.3. *Advances in Non-Linear Systems*

What better wish could there be for the future than that we may at least have a generalized non-linear theory? That would at least put us on the same footing as nature's servo system and allow us to get away from the false world of linear devices. Perhaps statistics will be called upon to play their part here? At least it seems possible that with very non-linear systems, optimization or adaptive control can only be applied by using a digital computer to do a piece-wise approach to the ideal operating points.

10. Acknowledgments

The authors would like to pay particular tribute to Mr. R. J. Perdue, who is jointly responsible with Mr. James for the activities of the System Engineering Group, for his efforts in the preparation of this paper. Thanks are also due to Mr. W. T. Lee for his contribution to the design and application of the computer.

Manuscript first received 19th May 1960 and in final form on 6th September 1960 (Paper No. 601).

U.D.C. 621.382.333:621.396.822

Variation of L.F. Noise Figure of a Junction Transistor

By

S. DEB, D.Phil.

AND

A. N. DAW, M.Sc.†

Summary: Some results of measurement of l.f. noise component over a bandwidth of 300 c/s centred about a mid-band frequency channel of 1000 c/s are reported for a number of commercial transistors. Investigations cover a range of collector current (I_c) 0·3–1·6 mA and a temperature range −20° to +45°C. It is found that the noise increases with increasing values of I_c and decreasing values of temperature. Simultaneous measurement of lifetime τ of minority carriers in the base region shows that both $1/\tau$ and the l.f. noise vary in the same manner with temperature. Results are discussed in the light of the current ideas regarding the origin of l.f. noise in semiconductor junction. It is shown that the major part of such noise in the transistors investigated may be attributed to surface recombination noise. It is tentatively suggested that the spectral distribution function $\psi(f)$ of surface recombination noise varies exponentially as the surface recombination velocity. Further, the value of $\psi(f)$ should, in general, depend on the current I_c and when this is taken into account the noise should vary as the square of I_c. For a given I_c-value the l.f. noise figure depends very much on the functional relationship between $\psi(f)$ and I_c.

1. Introduction

Noise in transistors poses a problem of much greater complexity than that in conventional vacuum tubes. This is because apart from the random shot and thermal noise there also appears, at more than one region of transistors, the so-called flicker noise which behaves in a rather complicated manner. Further, the correlation between the noises generated at the various regions is not uniquely known. Of the different types of noises in transistors the shot and the thermal noise components have been studied most, possibly because of the emphasis laid on the improvement of the high frequency performance of transistors. The flicker noise component has not, however, been investigated in detail. Further, investigations[1–3] carried out so far on flicker noise are concerned more with its dependence on frequency—a question of universal interest in all types of flicker and contact noises. For a clear understanding of the origin of flicker noise, it is also of interest to know how it varies with (i) the input current and (ii) the temperature. But unfortunately only a limited amount of work[1, 4, 5] has so far been done on these problems. The purpose of the present paper is to communicate some results of noise measurement on transistors carried out in this laboratory with a view to determining these variations. The theory and method of experimental measurement are first described in Section 2. Results of measurement are given in Section 3 which is followed in Section 5 by a critical examination of these results and an attempt at tracing the possible origin of the observed l.f. noise in the transistors under investigation. For convenience of discussion, this is preceded in Section 4 by a short account of the present position of our knowledge about flicker noise.

2. Experimental Arrangement

Noise figure was measured by the method of injection of a small equivalent signal at the transistor input. In the experimental set-up, the noise output of the transistor was fed into a high-gain amplifier with maximum response at 1000 c/s and possessing a bandwidth of 300 c/s. The amplifier was terminated by a suitable power measuring device which gave a reading proportional to

$$4kTR_gFA^2\Delta f$$

where k = Boltzmann's constant,
T = absolute temperature,
R_g = source resistance,
F = noise figure of the transistor,
A = voltage gain of the system comprising the transistor and the amplifier that follows, and
Δf = the equivalent noise bandwidth of the system.

It is assumed here that the amplifier itself does not introduce any significant amount of noise. An audio

† Institute of Radio Physics and Electronics, University of Calcutta.

signal, V_e say, at a frequency of 1000 c/s was then introduced in series with the input resistance of the transistor, so that the reading of the output power meter became proportional to

$$(\overline{V_e^2}+4kTR_gF\Delta f)A^2$$

V_e was so adjusted that the power output as recorded by the measuring device was twice that obtained with the audio source turned off. Under this condition,

$$(\overline{V_e^2}+4kTR_gF\Delta f)A^2=2\times 4kTR_gFA^2\Delta f \qquad (1)$$

or

$$\overline{V_e^2} = 4kTR_gF\Delta f$$

or

$$F = \frac{\overline{V_e^2}}{4kTR_g\Delta f} \qquad \ldots\ldots\ldots(2)$$

Eqn. (2) provides a simple relation for the measurement of F if Δf is measured separately.

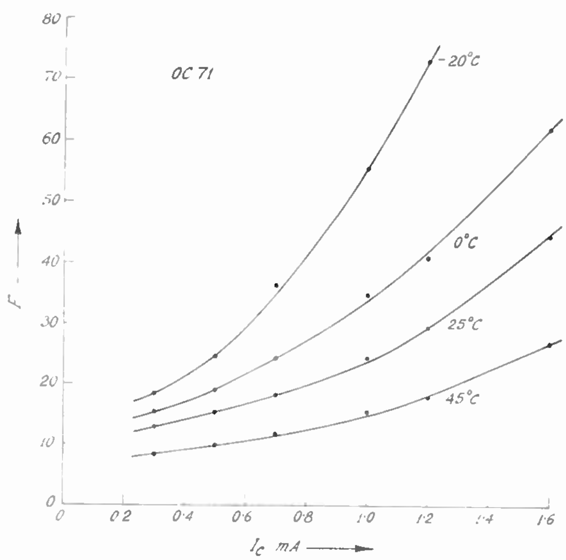


Fig. 1. Variation of the noise figure with collector current for transistor type OC71 at different operating temperatures.

Observations were taken on three different types of commercial transistors, namely, OC71, 2N206 and 2N220. Readings were taken over a range of input current corresponding to collector current values from 0·3 mA to 1·6 mA and at four different temperatures, 45°, 25°, 0° and −20° C.

It is to be noted that the noise figure as determined by the above method is the overall value comprised of both the flicker and shot noise. Of these the contribution of shot noise to the noise figure is given by [6]

$$F_{\text{shot}} = \frac{r_b'}{R_g} + \frac{r_e}{2R_g} + \frac{(r_e+r_b'+R_g)(1-\alpha)}{2r_eR_g\alpha} \qquad \ldots\ldots(3)$$

where r_b' = extrinsic base resistance,
r_e = emitter junction resistance and
α = current amplification factor.

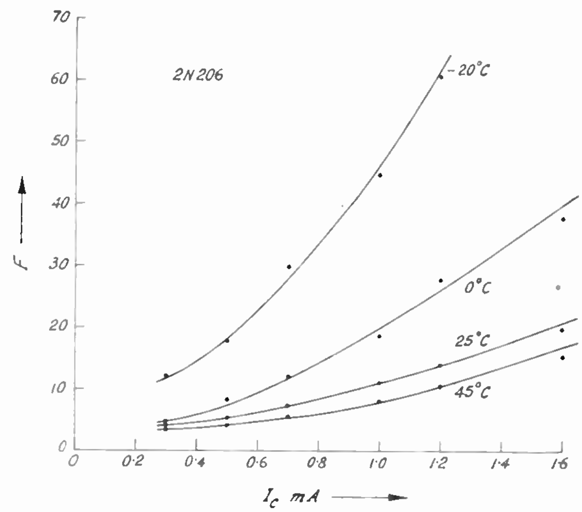


Fig. 2. Variation of the noise figure with collector current for transistor type 2N206 at different operating temperatures.

Thus F_{shot} can be determined if the parameters r'_b, r_e, R_g and α are measured. Subtracting this from (2), one gets the true contribution due to flicker noise. It may be noted that in a small frequency range around 1000 c/s, the term F_{shot} is usually very small in comparison with (2).

3. Experimental Results

Observations were taken for all three transistor types with the transistors connected in the common-base mode. The values of noise figure to be reported in this section are those obtained after subtracting the contributions due to shot noise and may be taken as a measure of the effect due only to the flicker noise. In what follows the term noise figure will be used to signify this residual value arising out of the low frequency component of noise alone.

Values of noise figures for transistor types OC71, 2N206 and 2N220 as a function of the collector current

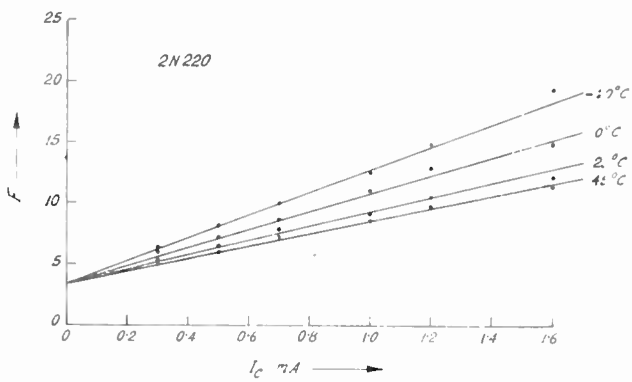


Fig. 3. Variation of the noise figure with collector current for transistor type 2N220 at different operating temperatures.

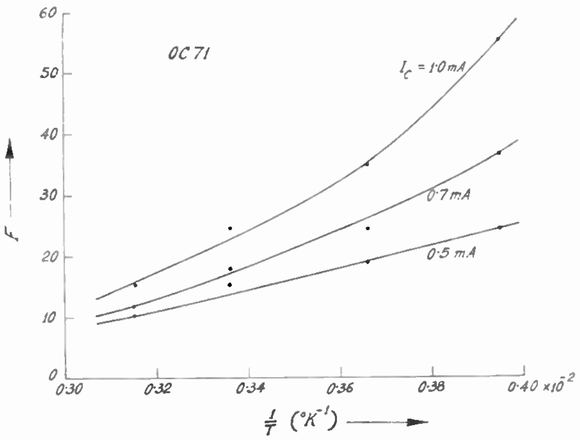


Fig. 4. Variation of the noise figure with operating temperature for transistor type OC71 at different values of collector current.

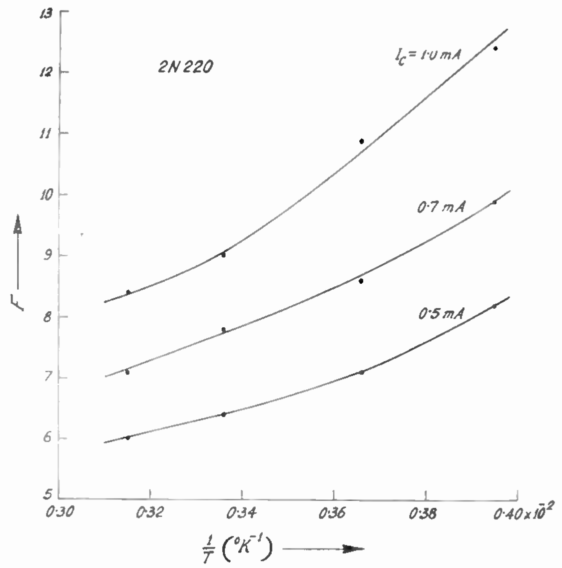


Fig. 6. Variation of the noise figure with operating temperature for transistor type 2N220 at different values of collector current.

I_c with temperature T as parameter are plotted in Figs. 1, 2 and 3. The same values are presented in Figs. 4, 5 and 6 as a function of temperature with I_c as parameter. It is observed from these figures that in general the noise figures for all the three transistors increase with increase in I_c and decrease in temperature. The precise laws of variation are, however, different for different types of transistors. The variation is found to be linear with I_c for transistor type 2N220 (Fig. 3) and super-linear for the other two (Figs. 7 and 8). It may also be pointed out that the values of noise figure do not become unity at $I_c = 0$, a condition under which transistor action ceases. The noise figure may, therefore, be regarded as being composed of two parts: (i) a part δ_1 independent of I_c but dependent on T; (ii) a part δ_2 dependent on both I_c and T. Examining these two components separately one finds that for transistor types OC71 and 2N206, both δ_1 and δ_2 decrease with increase in temperature. For the type 2N220, δ_1 is almost independent of temperature while δ_2 varies in the same manner as for the other two types although the variation with I_c is comparatively less rapid. It should be noted that type 2N220 is classified under the so-called low-noise family of transistors. The above results indicate that its low noise is due both to a low value of δ_1 and to a relatively slower variation of δ_2 with I_c. From the point of view of the actual values of noise figure the

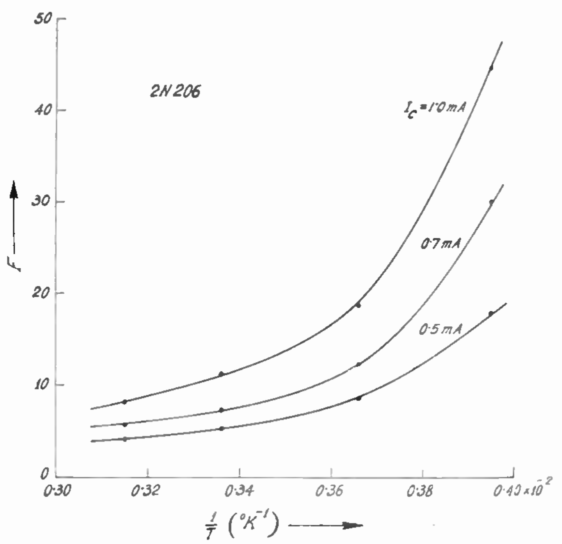


Fig. 5. Variation of the noise figure with operating temperature for transistor type 2N206 at different values of collector current.

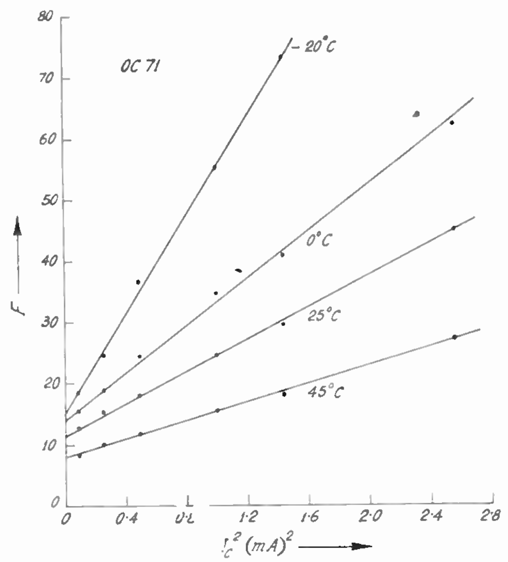


Fig. 7. Noise figure as a function of the square of the collector current for transistor type OC71 at different operating temperatures.

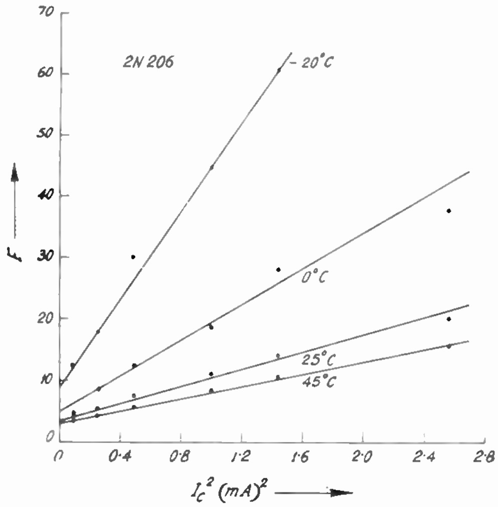


Fig. 8. Noise figure as a function of the square of the collector current for transistor type 2N206 at different operating temperatures.

difference between types 2N220 and 2N206 is not marked. But the general trends of variation of noise figure of the two types have little in common. On the contrary, the variation for type 2N206 resembles more that of type OC71.

4. Current Ideas regarding the Probable Source of Origin of L.F. Noise and Expressions for Noise Figure

Before any comment can be made on the results mentioned in the preceding section it will be helpful to recall in brief the present views[7] on the origin of l.f. noise in an alloy junction transistor and to introduce, on the basis of these views, a theoretical relation giving the noise figure of the device. L.f. noise can originate at both the emitter-base and the collector-base junctions. It is believed that for a given junction the noise might originate as either surface noise as an accompaniment of the recombination process taking place at the region or leakage noise owing to the flow of a leakage current across the junction. Leakage noise is characteristically prominent for an ill-formed junction and when the voltage drop across the junction is large—greater than, say, 10 volts. As such it is large in a reverse biased junction and, for a given bias voltage, larger for a junction having higher leakage current. Surface noise is a function of the current flowing and the so-called surface recombination velocity. In transistors under the usual operating conditions, the effects of these noise sources are further enhanced by the base conductivity modulation. Taking all these into account the transistor equivalent circuit in the common base mode along with the noise generators is usually represented as in Fig. 9. The expression for noise figure derived for the general case is quite complex. But simplification is possible owing to the fact that under usual operating conditions of a transistor amplifier the emitter-base junction is forward biased with little voltage drop across the junction and the collector-base junction is reverse biased to several volts with little surface recombination taking place near it. It is, therefore, permissible to retain only the surface noise generator at the former junction and the leakage noise generator at the latter. The equivalent circuit then takes the form as shown in Fig. 10. On the basis of this circuit one can, by well-known procedure,[1, 8, 9] derive the following expression for the noise figure of a common base junction transistor:

$$F = 1+\frac{(r_b'+R_g)^2}{4kTR_g}\left[1-\frac{I_b}{(R_g+r_b')}\cdot\frac{\partial R_b'}{\partial I_b}\right]\times$$
$$\times\left[\overline{i_{s1}^2}\oplus\overline{i_{L2}^2}\left\{\frac{1}{\alpha}\left(1+\frac{r_e}{r_b'+R_g}\right)\right\}^2\right] \quad\ldots\ldots(4)$$

where I_b = base current,

R'_b = d.c. base lead resistance and is related to the a.c. base lead resistance r'_b by

$$r_b' = R_b' + I_b\frac{\partial R_b'}{\partial I_b} \quad\ldots\ldots\ldots(5)$$

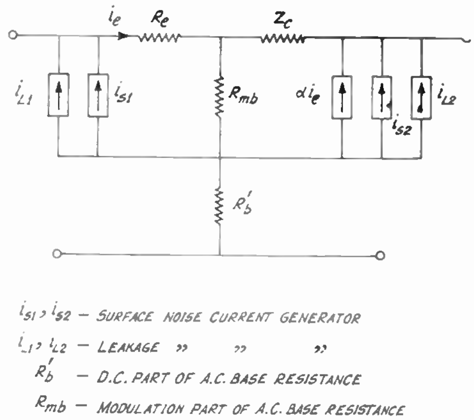


Fig. 9. Equivalent l.f. noise circuit of an alloy junction transistor

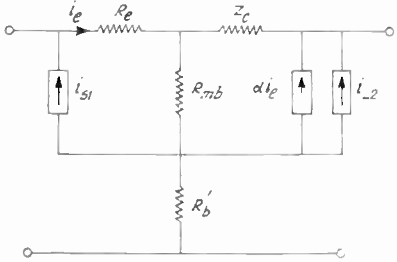


Fig. 10. Simplified representation of Fig. 9.

$\overline{i_{s1}^2}$ = mean square surface noise current originating at the emitter-base junction, and

$\overline{i_{L2}^2}$ = mean square leakage noise current originating at the collector-base junction.

It may be pointed out that the term $I_b(\partial R'_b/\partial I_b)$ on the right-hand side of eqn. (4) represents the contribution due to base conductivity modulation.

The exact expressions for the equivalent noise generators $\overline{i_{s1}^2}$ and $\overline{i_{L2}^2}$ are still matters of speculation. Fonger,[1] however, gives the following relations:

$$\overline{i_{s1}^2} = \frac{2(I_c - I_{cs} - I_{ec} - I_L)^2}{\pi^2 D^2}\left(\frac{W}{R_e}\right)^3 \psi(f)\Delta f \quad \ldots\ldots(6)$$

$$\overline{i_{L2}^2} = 16kT \times 10^{13} I_L^2 \frac{\Delta f}{f} \quad \ldots\ldots(7)$$

where I_c = the collector-base bias current,
I_{cs} = the collector-base saturation current,
I_{ec} = the emitter-collector saturation current,
I_L = the collector-base leakage current,
W = the width of the base region,
D = the diffusion constant of minority carriers in the base region,
R_e = the radius of the emitter dot, and
$\psi(f)$ = the so-called spectral distribution function characterizing the noisiness of the surface.

Explicit expressions for the frequency dependent term $\psi(f)$ are not known. It varies roughly as $1/f$. Further, Fonger suggests that $\psi(f)$ decreases with increase in injection level and remembering that I_c is very large compared to I_{cs}, I_{ec} and I_L, the exponent of I_c in the expression for the noise figure might vary within the range 2 to 1. It is also reasonable to expect that $\psi(f)$ should be a function of the surface recombination velocity s. A general correlation between the two does not appear to have been established although it is known that both can be reduced by proper surface treatment.

Noise generator $\overline{i_{L2}^2}$ increases in proportion with the leakage current through the junction and with increase in the value of the collector bias in a rather complicated way. It also increases sharply with increase in temperature and is very sensitive to the surrounding atmosphere.

Bearing the above picture in mind, we proceed to discuss the experimental results described in Sect. 3.

5. Discussion of Results

As already found from an examination of Figs. 4 to 6 the observed noise figure decreases with increase in temperature indicating that the leakage noise component can at best be small in the transistors tested. This conclusion was further supported by the facts that the observed noise figure was insensitive to changes in the collector bias voltage and that the leakage current was indeed quite small for the collector-base junction with the emitter end open.

Again, as shown in the Appendix, the conductivity modulation terms for all the three transistors under measurement were small compared to $(R_g + r'_b)$. Hence, disregarding the leakage noise term

and putting $$I_b \frac{\partial R'_b}{\partial I_b} \ll (R_g + r'_b)$$

eqn. (4) may be written as

$$F = 1 + \frac{2(r'_b + R_g)^2}{4\pi^2 D^2 kTR_g}\left(\frac{W}{R_e}\right)^3 I_c^2 \psi(f) \quad \ldots\ldots\ldots(8)$$

According to this expression, for $I_c = 0$, $F = 1$ whereas from experimental results for $I_c = 0$, $F = \delta_1 > 1$. The observed results can, therefore, be accounted for by adopting either of the following two alternative views:

(i) the component δ_2 of the observed noise figure is due to surface noise and the component δ_1, independent of I_c, is due to the existence of additional noise sources not covered by the previous discussions, or

(ii) the whole of the observed noise figure is due to surface effects alone. This requires that $\psi(f)$ varies with I_c in a particular manner.

If the first of the above views is correct then the component δ_2 of the observed noise figure should behave in the manner required by the second term on the right-hand side of eqn. (8). Equating this latter term with δ_2 one obtains

$$\delta_2 = \frac{2(r'_b + R_g)^2}{4\pi^2 D^2 kTR_g}\left(\frac{W}{R_e}\right)^3 I_c^2 \psi(f) \quad \ldots\ldots\ldots(9)$$

Unfortunately, the expression for δ_2 as given by eqn. (9) includes the factor $\psi(f)$ which has not yet been explicitly determined. To isolate this factor, we rearrange eqn. (9) in the form

$$\frac{2}{\pi^2 W R_e^3} I_c^2 \psi(f) = \frac{\delta_2}{\dfrac{(r'_b + R_g)^2}{4kTR_g}\left(\dfrac{W^2}{D}\right)^2} = \delta_3, \quad \text{say} \ldots\ldots(10)$$

The right-hand side of eqn. (10) can be determined experimentally by known procedure. In Figs. 11, 12 and 13, the experimental values of δ_3 for the three transistors are plotted as functions of I_c^2 for transistor types OC71 and 2N206 and of $I_c^{3/2}$ for transistor type 2N220. It is seen that for the first two transistors, δ_3 behaves in the same manner as required by eqn. (8). Corresponding variations for transistor type 2N220 can be theoretically accounted for only if it is assumed that for this transistor $\psi(f)$ is current dependent and includes a factor $I_c^{-1/2}$. As already stated the possibility of such a dependence has been pointed out by Fonger.

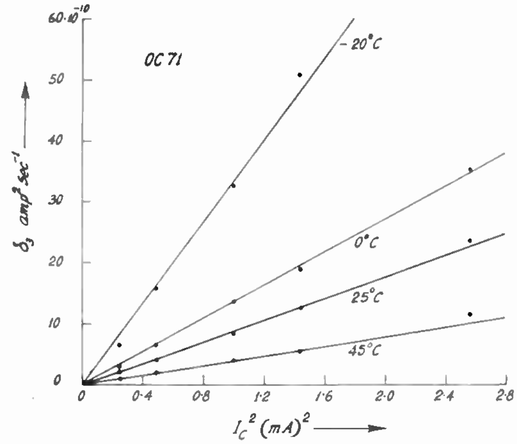


Fig. 11. Showing δ_3 [see eqn. (10)] as a function of I_c^2 for transistor type OC71.

Let us now consider the temperature dependence of δ_2. Experimental results show that both δ_2 and δ_3 are strongly temperature dependent. In fact, δ_2 and δ_3 are found to satisfy the following empirical relations:

$$\delta_2 = CI_c^m \exp\left(\frac{E}{T}\right) \qquad \text{.........(11)}$$

$$\delta_3 = PI_c^n \exp\left(\frac{Q}{T}\right) \qquad \text{.........(12)}$$

The constants C, E, P, Q, m, and n vary with the type of transistor. Their values for the three types of transistors investigated, are given in Table 1.

Table 1

Transistor type	C	E	P	Q	m	n
OC71	$9{\cdot}27\times10^{3}$	$2{\cdot}13\times10^{3}$	$1{\cdot}15\times10^{-7}$	$2{\cdot}60\times10^{3}$	2	2
2N206	$6{\cdot}76\times10^{3}$	$2{\cdot}77\times10^{3}$	$2{\cdot}65\times10^{-8}$	$3{\cdot}00\times10^{3}$	2	2
2N220	$3{\cdot}45\times10^{3}$	$0{\cdot}85\times10^{3}$	$5{\cdot}60\times10^{-7}$	$1{\cdot}18\times10^{3}$	1	$1\frac{1}{2}$

Equations (10) and (12) suggest that a part at least of the observed variation with temperature may be due to a dependence of $\psi(f)$ on T. Recalling eqns. (9) and (10) we find that in order to determine explicitly the variation of $\psi(f)$ with temperature a study must be made of the entity δ_3/I_c^2 which gives $\psi(f)$ apart from a scaling factor $2/WR_e^3\pi^2$. From eqn. (12) and Table 1, we, however, find that such a study leads to the empirical relation

$$\frac{\delta_3}{I_c^2} \propto \exp\left(\frac{Q}{T}\right) \qquad \text{.........(13)}$$

showing that $\psi(f)$ should vary exponentially with temperature. It may, therefore, be concluded that the observed temperature dependence of δ_2 is due mainly to the variation of $\psi(f)$ with temperature. Now, as pointed out already, $\psi(f)$ should be a function of the surface recombination velocity and, for transistors in which recombination occurs mainly on the surface, also of the lifetime of minority carriers in the base region. Experimental results obtained by other workers[10-13] show that the surface recombination velocity s and the lifetime τ are indeed sensitive to temperature variation. In the present case, for all the three transistors, the lifetime τ was found to increase slowly with increase in emitter current, showing that recombination at the surface must have been the dominant process.[12] In order to test whether variations of τ and $\psi(f)$ are correlated, values of the former were measured by determining the cut-off frequencies in the common emitter mode. In Fig. 14, values of $\ln\frac{2}{WR_e^3\pi^2}\psi(f)$ and $1/\tau$ are plotted as functions of $1/T$ for all the three transistors. It is observed from this figure that both $\ln\frac{2}{WR_e^3\pi^2}\psi(f)$ and $1/\tau$ vary linearly with $1/T$. While too much significance cannot be attached to these findings in view of the limited range of observation, it is interesting to point out that the observed dependence of noise figure on temperature may be explained if it is assumed that the spectral distribution function $\psi(f)$ varies exponentially as $1/\tau$ or, if the recombination rate is not too large,[14] as the surface recombination velocity s. It is to be noted that such a dependence had also been anticipated earlier.

Let us now discuss the probable origin of δ_1. Two sources may be mentioned: (i) leakage noise of the collector-base junction and (ii) contact noise or additional noise sources which appear across the

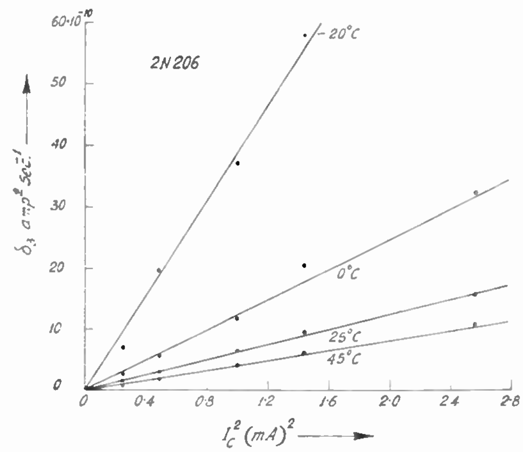


Fig. 12. Showing δ_3 [see eqn. (10)] as a function of I_c^2 for transistor type 2N206.

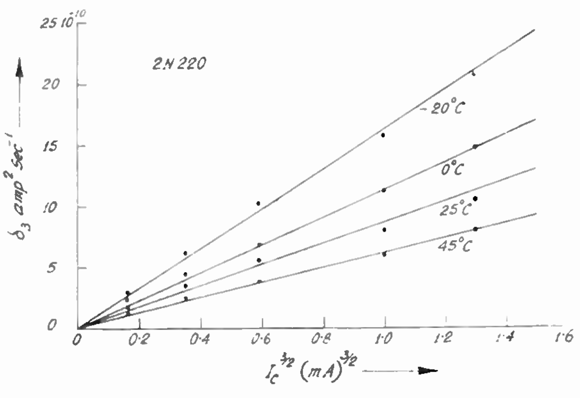


Fig. 13. Showing δ_3 [see eqn. (10)] as a function of $I_c^{3/2}$ for transistor type 2N220.

emitter-base junction on application of the bias voltage owing to the imperfect nature of the junction.[5] As already mentioned, leakage noise was negligible for the transistors investigated under the usual operating conditions. More importance should, therefore, be given to the second hypothesis, namely, additional noise sources arising out of a badly-formed junction.

We now proceed to discuss the observed noise figure on the basis of the second view mentioned earlier, that the observed noise is solely due to surface effects. This obviously requires that for the transistors under investigation $\psi(f)$ should be expressible in the form

$$\psi(f) = A + \frac{B}{I_c} + \frac{C}{I_c^2} \qquad \text{.........(14)}$$

Combining this with eqn. (8), the expression for the noise figure of a transistor may be written as

$$F = 1+A'+B'I_c+C'I_c^2 \qquad \text{.........(15)}$$

where

$$\left.\begin{aligned} A' &= \frac{C(r_b'+R_g)^2}{2\pi^2D^2kTR_g}\left(\frac{W}{R_e}\right)^3 \\ B' &= \frac{B(r_b'+R_g)^2}{2\pi^2D^2kTR_g}\left(\frac{W}{R_e}\right)^3 \\ \text{and}\quad C' &= \frac{A(r_b'+R_g)^2}{2\pi^2D^2kTR_g}\left(\frac{W}{R_e}\right)^3 \end{aligned}\right\} \qquad \text{.........(16)}$$

The coefficients A', B' and C' may be functions of temperature and may also depend on I_c through r'_b and D although this latter dependence is not very marked except for type 2N220. From Figs. 3, 7 and 8, and eqn. (15), the values of the constants A', B', and C' at 25° C are obtained as shown in Table 2. As regards the temperature dependence the observed results can be explained on the assumption that the constants A', B' and C' are functions of the surface recombination velocity. Obviously, this second point of view is simpler than the one discussed earlier.

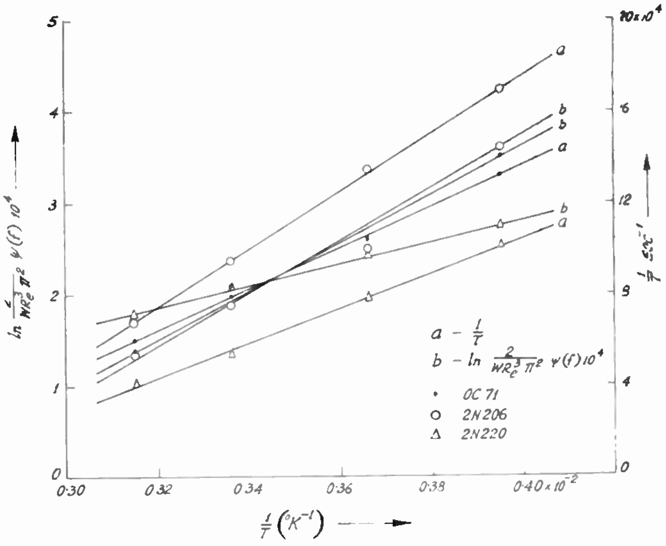


Fig. 14. Showing the variations of $\ln\frac{2}{WR_e^3\pi^2}\psi(f)$ [see eqn. (10)] and $1/\tau$ with temperature for transistor types OC71, 2N206 and 2N220.

Table 2

Transistor type	A'	B'	C'
OC71	10·5	0	13×10^6
2N206	2·5	0	7×10^6
2N220	2·5	$5{\cdot}75\times10^3$	0

6. Conclusion

The observed experimental results indicate that the low frequency noise in transistors may arise out of either a combination of surface effects and the imperfectness of junctions or on surface effects alone. The latter possibility no doubt gives a much simpler picture and is also quite adequate if $\delta_1 < \delta_2$—a condition which is satisfied for normal operating values of I_c. Irrespective of the correctness or otherwise of one or the other of these two points of view there are, however, several aspects common to both. For example, both require that $\psi(f)$ should be current dependent although the law giving this dependence is not identical in the two cases. Further, the temperature dependence of the constants A', B' and C' shows that $\psi(f)$ should also be temperature dependent—a conclusion in agreement with that arrived at from the first explanation. In fact, it follows from eqn. (15) and the expression for the noise figure considered earlier that the term $(1+A')$ can be identified with δ_1 and the remaining terms with δ_2. On this basis we find that the low noise in transistor type 2N220 is due both to a small δ_1 value and a simpler law of variation of δ_2 with I_c. Simultaneous measurement of lifetime and noise also show that the observed dependence of δ_2 on temperature may be explained if one assumes that the spectral distribution function $\psi(f)$ varies exponentially as $1/\tau$, or, if the recombination rate is not too large, as the surface recombination velocity s.

7. Acknowledgments

The authors would like to thank Professor S. K. Mitra, F.R.S., for his kind interest in the work. Thanks are also due to Professor J. N. Bhar for his constant encouragement during the progress of the work and helpful suggestions for improvement of the text.

8. References

1. W. H. Fonger, "A determination of $1/f$ noise sources in semi-conductor diodes and triodes", "Transistors I", p. 239 (R.C.A., Princeton, 1956).
2. H. C. Montgomery, "Transistor noise in circuit applications", *Proc. Inst. Radio Engrs*, **40**, p. 1461, 1952.
3. P. M. Bargellini and M. B. Herscher, "Investigations of noise in audio frequency amplifiers using junction transistors", *Proc. Inst. Radio Engrs*, **43**, p. 217, 1955.
4. K. Amakasu and M. Asano, "Temperature dependence of flicker noise in *n-p-n* junction transistors", *J. Appl. Phys.*, **27**, p. 1249, 1956.
5. C. A. Lee and G. Kaminsky, "Investigation of the temperature variation of noise in diode and transistor structures", *J. Appl. Phys.*, **30**, p. 1849, 1959.
6. E. G. Nielsen, "Behaviour of noise figure in junction transistors", *Proc. Inst. Radio Engrs*, **45**, p. 957, 1957.
7. A. van der Ziel, "Noise in junction transistors", *Proc. Inst. Radio Engrs*, **46**, p. 1019, 1958.
8. B. L. A. Wilson, "Transistor noise, its origin, measurement and behaviour", *J.Brit.I.R.E.*, **18**, p. 207, 1958.
9. D. Dewitt and A. L. Rossoff, "Transistor Electronics", (McGraw-Hill, New York, 1957).
10. W. N. Reynolds, "Surface recombination in germanium", *Proc. Phys. Soc.*, **B66**, p. 899, 1953.
11. Y. Kanai, "Temperature dependance of the surface recombination velocity in germanium", *J. Phys. Soc. Japan*, **9**, p. 292, 1954.
12. S. Deb and A. N. Daw, "On the lifetime and diffusion constant of the injected carriers and the emitter efficiency of a junction transistor", *J. Electronics and Control*, **5**, p. 514, 1958.
13. D. M. Evans, "Measurements on alloy type germanium transistors and their relation to theory", *J. Electronics*, **1**, p. 461, 1956.
14. W. Shockley, "Electrons and Holes in Semiconductors," (D. Van Nostrand, New York, 1950).

9. Appendix

The magnitude of the conductivity modulation term $I_b(\delta R'_b/\delta I_b)$ can be evaluated easily with the help of eqn. (5). The procedure consists in the measurement of r'_b for various values of I_b and plotting the same against I_b. As eqn. (5) evidently shows, a numerical integration is required to obtain the value of R'_b. Subtracting the value of R'_b thus determined from the measured value of r'_b, one gets readily the magnitude of the conductivity modulation term.

Proceeding in this way, values of the term $I_b(\delta R'_b/\delta I_b)$ for the transistors under investigation were calculated and found to be -63, -57 and -55 ohms for transistor types OC71, 2N206 and 2N220 respectively. Comparing these values with that of the term $(R_g+r'_b)$ for which typical values were of the order of 650 ohms, it was found that for the transistors under measurement the latter quantity was always much larger than the former. Accordingly, the contribution of the conductivity modulation term to the expression for the noise figure could be neglected without introducing any serious error in calculation.

Manuscript received 15th August 1960 (Paper No. 602).

U.D.C. 621.398

Engineering Aspects of Missile Telemetry Equipment—

The Airborne Sender for 24-channel Telemetry

By

W. M. RAE, B.SC.†

Presented at a Symposium on Radio Telemetry held in London on 25th March 1959.

Summary: After a brief description of the 24-channel a.m. f.m. telemetry system, some of the design problems arising with the airborne sender as described. Production problems are discussed in relation to certain of the sub-assemblies. Finally mechanical, electrical and aerial design aspects of adapting the sender for use with a typical missile are considered.

1. Introduction

The monitoring of missiles in flight is mainly concerned with the behaviour of control and guidance systems. These are essentially servo systems and are unlikely to have a frequency response above 10 c/s. It is therefore necessary to use a telemetry system which will monitor a large number of low-frequency signals. Time-division multiplex is well suited to this requirement.

This paper provides a brief survey of typical design and development problems involved in producing a missile-borne telemetry sender employing time-division multiplex (t.d.m.).

2. Brief Description of Sender

It is usually necessary to transform the missile characteristics into a form suitable for telemetry. The method used varies from simple potential-divider circuits to special "pick-off" (or transducer) units, depending on the nature of the characteristics.

The inputs are connected to a 24-segment multiplexer switch via low-pass filters and a signal conversion unit. The low-pass filters are only necessary if it is possible for conditions in the missile circuits to cause the frequency of any of the inputs to increase to more than half the frequency at which each input is sampled by the multiplexing switch. The signal conversion unit matches the input channels to the modulator which, as shown later, reduces errors. On the 24 inputs, one segment is normally reserved for synchronization purposes and at least two others for in-flight calibration (Fig. 1).

† E.M.I. Electronics Ltd., Telemetry Division, Feltham, Middlesex.

The output of the modulator is a frequency modulated signal varying in frequency over the range 130–160 kc/s for input signals of ± 3 volts. The sub-carrier rises to a higher frequency (commonly 186 kc/s) during the synchronization interval. The frequency-modulated wave amplitude-modulates an r.f. oscillator beyond 100%. Consequently the r.f. oscillator is switched on and off by the sub-carrier waveform. Typical waveforms occurring during three channels are shown in Fig. 2.

3. Typical Design Problems

Much of the theory necessary to determine telemetry parameters (such as sub-carrier frequency and deviation) has already been published. However, some of the more important sub-unit performance requirements are briefly outlined.

3.1. *Limiting the Frequency of Input Signals*

If the spectrum of the sampled waveform exceeds half the sampling frequency the high frequencies are heterodyned and appear together with the low frequencies at the low end of the spectrum. Consideration of a simple case will illustrate the mechanics behind the general theorem. If a sine wave of frequency f_1 is sampled at frequency f_s and if $f_1 = f_s$ then the output as indicated by the envelope of the sampling pulses is d.c. The d.c. output level will be zero when the sampling point is at the cross-over of the sine wave and the d.c. output level will change between limits corresponding to the positive and negative excursions of the sine wave if the phase of the sampling pulses is varied. If the sine wave frequency f_s is increased so that the interval between samples is, say, one and a quarter times the period of the sine wave, then the point at which the sine wave is sampled changes by a quarter of a cycle at each successive sample. The envelope of the sampling pulses indicates a sine wave of amplitude equal to the input sine wave but at one fifth of the input frequency. The heterodyned frequencies cannot be eliminated, consequently an incorrect waveform appears at the receiver.

In a practical telemetry system it is not possible to obtain perfect frequency response up to half the sampling frequency and perfect cut-off above half the sampling frequency. Because of this the information bandwidth of a t.d.m. system will usually be limited to frequencies not greater than ⅓ or ¼ of the sampling

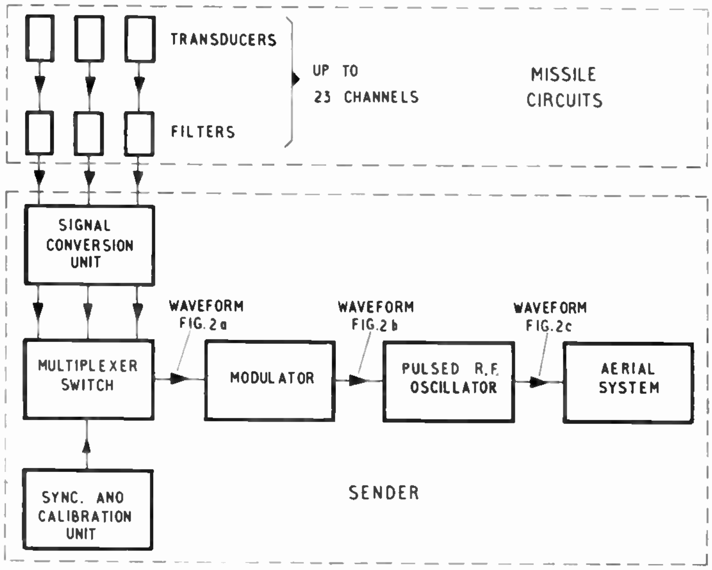


Fig. 1. Block diagram of 24-channel t.d.m. telemetry sender.

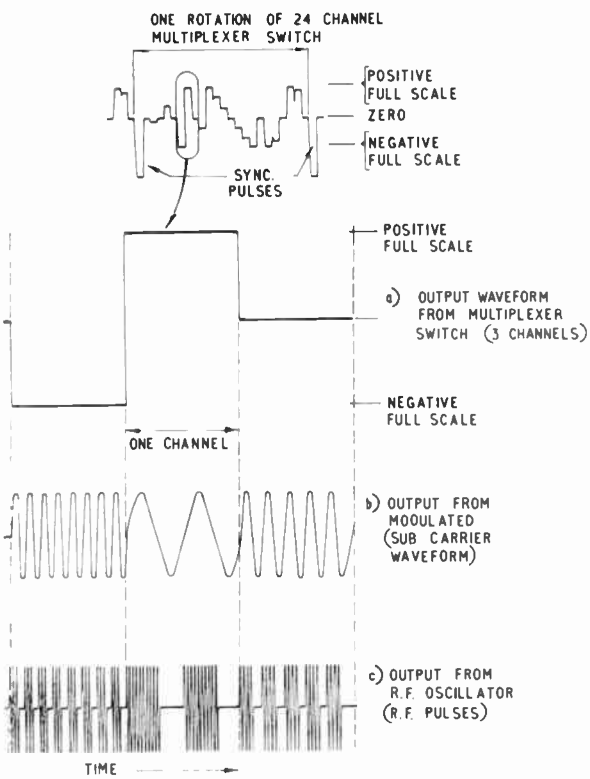


Fig. 2. Basic sender waveforms.

frequency. Input filters may be necessary to ensure that this condition is realized.

3.2. *Matching the Missile Circuits to the Sender*

To avoid pick-up on the leads from the missile circuits it is customary to incorporate a decoupling capacitor at the sender. This capacitor together with the transducer source resistance can introduce large errors. The equivalent circuit is shown in Fig. 3(a): R1 is the transducer resistance and C1 is the decoupling capacitor. The multiplexer switch connects this transducer to the modulator for a short period during each revolution. The equivalent input circuit of the modulator is a resistance and capacitance in parallel (R2 and C2 respectively).

The waveforms shown in Figs. 3(b), (c) and (d) show two main effects. During the sampling period (when the switch is closed) the value of the transducer voltage will differ from the open-circuit value, and, in addition, may vary during the period, e.g. as in Figs. 3(b) and (c). The fundamental reason for the variation in indicated value during the sampling period is the presence of CR networks on both sides of the switch causing a finite time to elapse between the instant of closing the switch and the achievement of steady state conditions. The waveforms of Figs. 3(b), (c) and (d) show the nature of these variations when the transducer time constant is less than, equal to, and greater than the modulator input time constant. The receiver gating interval is of shorter duration than the sampling interval and the inevitable variations in the timing of the receiver gating interval cause errors since the input waveform is also varying with time in the manner described. The potential dividing effect caused by the switch connecting two CR networks could be taken into account in the calibration but there is an associated cross-talk effect which cannot be allowed for. This cross-talk would not be present if the potential on C2 of Fig. 3(a) always had the same value just before the switch connects a transducer or input. In practice C2 acquires a different charge during each sampling interval and does not revert to a standard condition during the short open circuit period between sampling intervals. Hence the value of one input voltage influences the indicated value of succeeding input voltages.

3.3. *Errors introduced by the Switch*

Change in speed due to missile acceleration forces will result in a change in channel duration whereas the interval between receiver demultiplexing strobes remain the same. The strobes will therefore shift in position relative to the channel and will introduce errors. The receiver does correct for long term changes in switch speed but since synchronizing pulses are produced only once per switch revolution, sampling rate changes which are rapid by comparison with the switch speed cannot be corrected in the receiver.

Dimensional tolerances on the switch produce the same effects as speed variations. These tolerances can produce two important defects: irregular positioning of the edges of the commutator segments, and eccentricity of the rotor with respect to the stator (or commutator). Both types of defect result in irregular sampling. The former gives rise to random irregularities in the arrival times of samples; the latter causes cyclic irregularities. High contact resistance in the switch introduces unwanted attenuation and any variation of contact resistance will result in signal fluctuations which cannot be subsequently corrected. When inductive-type transducers are used, the switch is in the oscillator tuned circuit so that a high contact-resistance will reduce the Q factor, cause spurious sub-carrier frequency changes, and might cause oscillation to cease altogether.

The generation of contact potentials in the switch can also cause an appreciable error. A full-scale deflection of 6 volts is normally used so that a contact potential of 60 mV (which is quite possible with certain combinations of stator and rotor materials) will cause an error of 1%.

If the capacitance between segments is greater than about one or two picofarads, quite severe adjacent-channel crosstalk can be introduced with inductive transducers. This is because the adjacent channel transducer is coupled into the tuned circuit via the inter-segment capacitance, thereby affecting the resonant frequency.

The extent of frequency change due to this cause increases with increase of inter-segment capacitance. It also increases with increase of Q factor of the transducer.

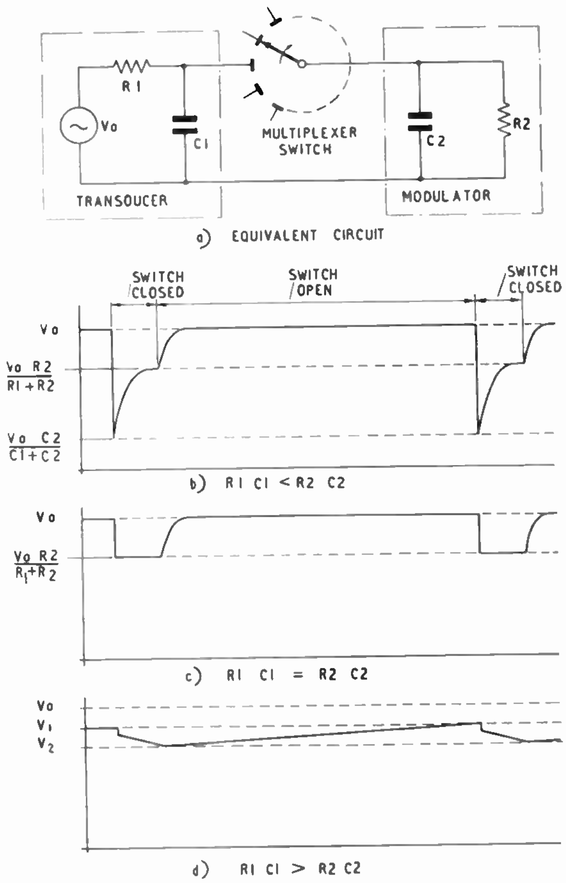


Fig. 3. Transducer/modulator matching showing the cyclic changes of potential on C1.

3.4. *Use of Calibration Signals to reduce Errors caused by Modulator Drift*

It is reasonably easy to provide a high order of linearity in the input voltage/sub-carrier-frequency output characteristic of the modulator. However, sub-carrier frequency drift with change in temperature, etc., causes appreciable error. It is usual, therefore, to allocate at least two channels to in-flight calibration. Using these calibrations channels as references, automatic correction for errors can be included in the receiving equipment.

3.5. *Errors introduced by the R.F. Oscillator*

Unless special precautions are taken, the frequency of an r.f. oscillator can change appreciably due to a number of causes. If these changes take place during the flight of a missile then the system performance will be degraded. The 24-channel system uses over-modulation rather than conventional amplitude modulation, a type of modulation often referred to as pulse operation. A disadvantage of this system is that noise occurring in the input circuit of the oscillator during the time it is switched off causes the leading edge of the pulse to be subjected to considerable phase jitter. This is undesirable because the phase jitter is still present in the detected output of the receiver and appears at the output of the sub-carrier filter as phase modulation. The ultimate effect is an unwanted fluctuation at the output. In bad cases the error due to this cause can be several per cent of full scale.

The mechanism of this leading edge jitter can be briefly explained as follows. When an oscillator is switched on, the amplitude of oscillation starts to grow exponentially. The rate of rise immediately after being switched on is restricted by the Q of the tuned circuit. It also depends upon the envelope of the noise voltage which happens to exist at the instant of switching. The smaller the noise voltage the slower the rise. When therefore the envelope of the noise is large at the instant of the switching, the pulse quickly grows to full amplitude. When it is small, a delay occurs before full amplitude is reached.

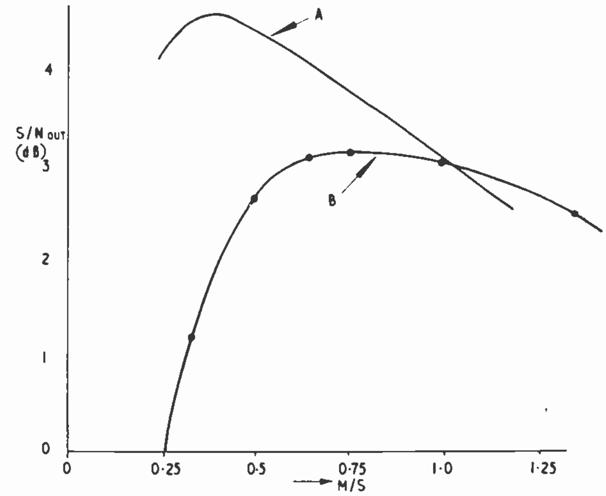
One way of overcoming this jitter is to inject a small signal into the oscillator tuned circuit during the interval between pulses. If this can be made substantially larger than the noise, then the oscillation will always commence from roughly the same point on the curve and hence jitter will be small.

This is done with the aid of a second very-low power oscillator coupled loosely into the tuned circuit of the main oscillator and tuned to half the transmission frequency. This secondary oscillator is called a "keep alive" for obvious reasons. When such an arrangement is used, noise on the system output due to pulse jitter is greatly reduced.

3.6. *Mark/Space Ratio*

During the pulse the noise is greater in amplitude and different in character from that which exists during the interval between pulses. If therefore the peak power is adjusted so as to keep the mean power constant as the mark/space ratio is varied, then though the input-signal/noise ratio is constant, the output-signal/noise ratio will depend upon the mark/space ratio.

Curves showing how the signal/noise ratio depends on the mark/space ratio are shown in Fig. 4. Two cases have been evaluated, one when the mean power is kept constant and the other when the peak power is constant. From these curves it can be seen that the optimum mark/space ratios are approximately 0·4 and 0·8 when mean power and peak power respectively are kept constant.



A Constant average input signal power

$$\frac{\text{average input signal power}}{\text{average input noise power}} = 2 = 3 \text{ db}$$

B Constant peak input signal power

$$\frac{\text{peak input signal power}}{\text{average input noise power}} = 4 = 6 \text{ db}$$

Fig. 4. Detector signal/noise ratio (S/N) output against mark/space ratio (M/S).

4. Development of Telemetry Sub-Assemblies

4.1. *Oscillators*

The importance of simplicity and small size in missile telemetry results in a preference for a single-valve oscillator for the r.f. transmitter. However the performance of such a simple device is only marginally better than the system performance requirements.

The oscillator uses a disc-seal triode in a Colpitts configuration (Fig. 5). The tuned circuit is connected between anode and grid with a capacitance divider giving feedback to the cathode. The tuned circuit consists of an inductance which resonates with the anode to grid capacitance. The inductance is provided by a short-circuited transmission line whose length is less than a quarter wavelength. A Lecher-line is preferred to a coaxial line because a given value of inductance can be obtained with a shorter length of line.

Since the system uses pulse modulation, and since modulation linearity is unimportant, grid modulation is preferred to anode modulation because a lower driving power is required. Pulse jitter is reduced by

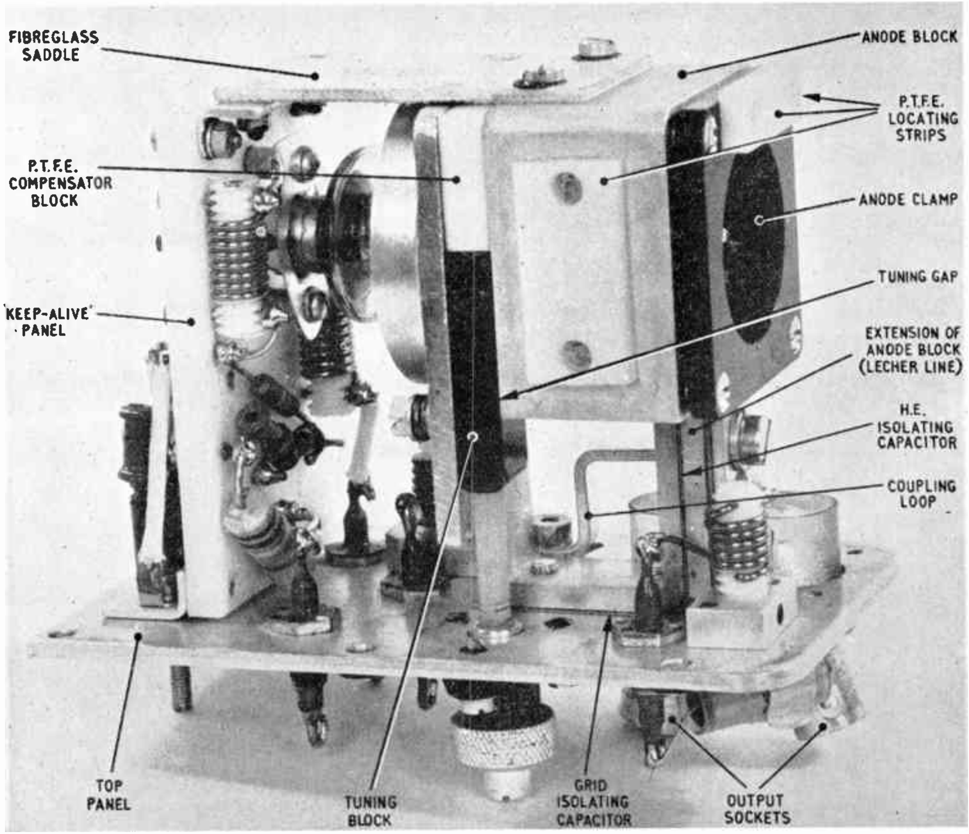


Fig. 5. Oscillator assembly.

the keep-alive oscillator which is loosely coupled to the tuned circuit.

At the completion of basic design this oscillator was found to suffer from two major electrical defects, namely a wide variation in frequency range and centre frequency between different oscillators, and excessive frequency drift with change in temperature.

The variation in frequency range was found to be due to differences in the width of the gap between the tuning and anode blocks. The main causes were dimensional tolerances on the p.t.f.e. compensating block and the tuning block. Extremely thin metal shims were provided for insertion between the p.t.f.e. block and the grid Lecher-line. These are added during assembly to allow the gap between the p.t.f.e. block and the anode block to be adjusted to 0·017 in. $\pm$0·002 in. This eliminated frequency-range variation.

The variation in centre frequency was found to be due to variation in the value of the h.t. isolating capacitor. This capacitor was formed by inserting a mica strip between the anode Lecher-line and the extension from the anode block. The capacitance therefore depended on the mechanical tolerances achieved in manufacture. With the initial design the capacitance varied from 50 to 90 pF. By using several strips of thinner mica it was found possible to select the strips to control the capacitance to within $\pm$5 pF.

The drift was found to be due to a change in the value of the h.t. isolating capacitor during operation. This was caused by "give" in the nylon clamping screws as the temperature increased—thus loosening the assembly. The nylon screws were therefore replaced by metal screws, a mica insulating strip, and a metal clamping plate. This reduced the drift but it still exceeded the specification at maximum temperature. A lengthy investigation showed that high-spots on the metal faces forming the capacitor caused the mica strip to distort, thus affecting the capacitance. This was eliminated by precision grinding both surfaces, thus bringing the drift within the performance specification.

The oscillator now met the electrical requirements but several weaknesses were noted on mechanical environmental tests. Under high acceleration forces the complete Lecher-line assembly moved relative to the top panel. This was due to "give" in the nylon screws holding the Lecher-line assembly to the top panel. A mica strip was used as the insulating material —this enabled the assembly to be clamped to the panel. P.t.f.e. strips were also attached to the anode block and the anode clamp; ideally these should just touch the inside of the oscillator case. This cannot be achieved in practice because of the build-up of dimensional tolerances on the different parts of the oscillator, but the possible movement is restricted to a very small amplitude.

The p.t.f.e. panel on which the disc-seal-triode base is mounted was increased in thickness and strapped to the Lecher-line assembly to stiffen the assembly.

Further modifications were necessary to improve minor weaknesses. The layout of the keep-alive circuit was altered to improve component positioning and fixing. The nylon string used to hold the keep-alive valve to the panel was replaced by a valve clip. The r.f. chokes, originally self-supporting, were mounted on formers rigidly fixed to the oscillator. Solder pins replaced the solder tags on the h.t.1 and G inputs.

This process produced an oscillator which met the main vibration and acceleration conditions expected in operational use.

4.2. *Multiplexing Switch*

The application of much effort to the design of telemetry switches on both sides of the Atlantic has failed to evolve a comprehensive design procedure. The following data indicate typical operating characteristics for a satisfactory switch:

Contact resistance less than 10Ω.
Contact potential less than 0·1 millivolt.
Inter-segment capacitance less than 1 pF.
Inter-segment insulation greater than 100 MΩ.

To ensure a constant sampling rate throughout the cycle it is necessary to maintain the angular position of the commutator segments within a few minutes of arc. With practical production methods this sets a lower limit of about 1 in. for the commutator diameter. In addition the switch motor must have a high rotor inertia to minimize fast speed changes. These and other considerations have led to a switch assembly which is about 4 in. long. The power consumption of switches is from 6 to 10 watts. In missile applications a life of about 100 hours is very satisfactory. Life is usually limited by the onset of random high contact resistance.

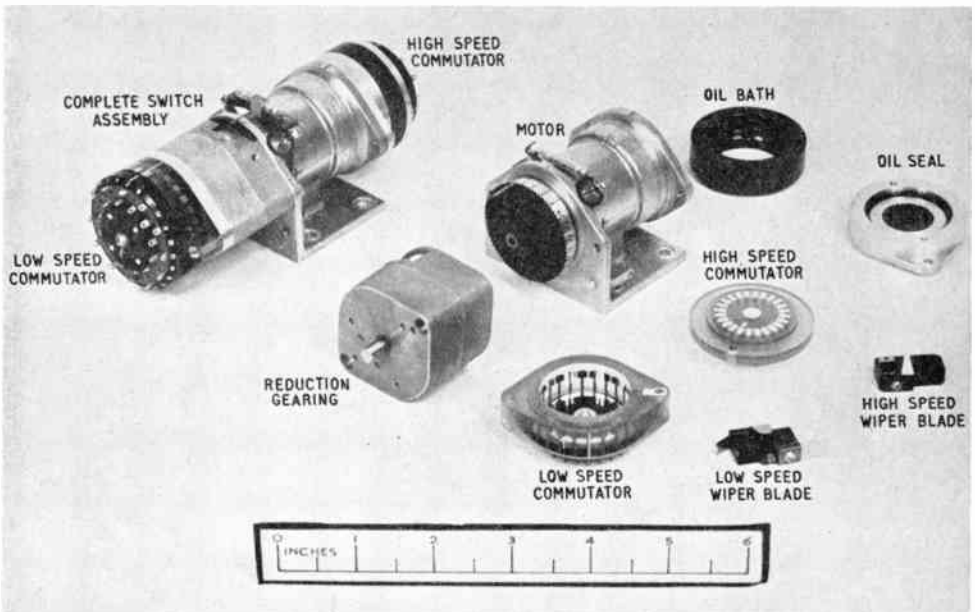


Fig. 6. Multiplexer switch assembly.

All these factors are dependent on the materials used in the construction of the switch. In this switch they are Araldite inter-segment insulation, brass segments and a nickel-silver wiper.

A double-ended multiplexer switch is shown in Fig. 6. This has a 24-channel high-speed commutator as described in the brief review of the sender. In addition, however, it has a low-speed commutator which can be used to telemeter h.t. potentials and other very slowly varying quantities.

The rate of wear is relatively high. However, this overcomes many problems because the light honing action continually rejuvenates the track. Despite the honing action track "poisoning" can still occur, especially when organic substances are present. This effect is still little understood. The most probable cause is that, under dynamic conditions a very thin film of the insulating material may be deposited on the working surface of the segments. On a molecular scale material can migrate to and fro between the wiper and the segments thus changing the nature of the working surfaces. The combined action of rubbing, and the making and breaking of electrical circuits can give rise to complex electro-chemical actions. Satisfactory operation can only be obtained when the switch contacts are immersed in oil. The most important function of the oil is believed to be that of keeping abrasive particles in suspension; this minimizes the build-up of conducting paths between the segments and also the effect of random interference of particles with the wiping action of the blade. One disadvantage is that an oil seal is necessary. This causes a large increase in motor load as well as adding directly to the size of the switch assembly.

Blade oscillation can cause partial or complete loss of contact between segment and blade. The blade constitutes a mass-spring system and will have several modes of oscillation. The blade must be designed to have a very low Q over the range of frequencies likely to be excited by the inevitable irregularities of the track and these irregularities must be kept to a minimum. Damping, which should be high, is provided mainly by blade pressure. However, the pressure must not cause the wear to become excessive: in practice a pressure of 100–120 grammes is used. The geometry of blade and commutator must be accurately controlled to ensure negligible changes in blade pressure. To avoid this the eccentricity between rotor and stator must be reduced to an extremely low value. Eccentricity as great as 0·003 in. would normally require piece-part tolerances as small as 0·0001 in. However, by using special jigs which enable critical selective adjustments during assembly, it has been possible to use piece-part tolerances of conventionally practical values.

4.3. *Airborne Power Supplies*

The design of the airborne power supply is normally complicated by limitations in space and weight, and by difficult environmental conditions.

Typical output requirements for the power supply are:

Modulator:	A stable h.t. of about 170 volts (20 mA). Bias for limiters, etc. (no current drain).
R.f. oscillator:	Unstabilized h.t. of up to 400 V (60 mA).
L.t.:	6·3 V for heaters and the multiplexing switch (up to 3·7 A).
Reference voltage:	Very stable low voltage (very low drain).

One of the most important supplies is the reference source. This is used for continuous in-flight calibration of the entire transmitter/receiver system. Some degradation of the absolute accuracy of the transmitter is permissible, with corresponding easing of power unit stability, provided the reference source is of a high-order of accuracy. Standard cells are not suitable for missile applications. However, the output voltage of mercury cells (Kalium) can be maintained to within 1% provided two main precautions are taken. These are to control the operating temperature to between say, 20–35° C and to limit the current drain to that required for a high-resistance potential-divider chain.

It is usual to supply the telemetry sender from an external power unit, normally mains derived, until just before the instant of launch. This eases the design of the missile-borne power unit since it is required to operate only during flight.

Telemetry supplies can be obtained either from the missile supply or from separate batteries. The latter system is normally preferred since it is completely independent of missile power unit failure, thus enabling the telemetry set to function unhampered when it is most required, i.e. when a fault occurs in the missile.

4.3.1. L.t. batteries

At the present time, one of the most suitable batteries for high-discharge-rate applications is the rechargeable, sintered-plate nickel-cadmium battery. The outstanding characteristic of the nickel-cadmium battery for missile applications is the high degree of reliability that can be achieved. The nickel-cadmium battery has a long life and can be frequently recharged or trickle-charged to maintain it in a state of readiness.

The battery will operate under severe conditions of shock, acceleration, and pressure and retains about 50% of its capacity after storage of one year at normal temperatures. If longer storage is required, the battery can be kept in a wet and discharged condition without deterioration and can be recharged readily to full capacity.

The capacity of the nickel-cadmium battery at high discharge rates is not unusual when compared with other available battery types, but its voltage regulation during discharge is extremely flat (typically a 5 ampere hour battery has an output of 1·3 V after charge, falling to 1·28 V after 4 min and to 1·2 V after 60 min at a current of 4 A). However, at low ambient temperatures, it is necessary to heat the battery to obtain satisfactory performance. This is usually accomplished just before the missile is launched—this ensures that the battery will stay warm during the relatively short flight.

4.3.2. H.t. supplies

These can be obtained either from the l.t. supply by rotary or electronic converters—or from a separate battery pack. The rotary converter is the least satisfactory since rotor and stator can foul under vibration.

Many factors are involved when choosing between an electronic converter and an h.t. battery pack. However, it is usually the watt-hour-capacity/volume ratio which is the deciding factor. When an electronic converter (usually transistorized) is used it is necessary to increase the capacity of the l.t. battery. Which system is used depends on whether the increase in l.t. battery volume plus the volume of the converter is greater than the volume of the h.t. battery.

5. Development of a Typical Missile Telemetry System

5.1. *Mechanical Design*

When designing telemetry equipment for a missile it is usually necessary to provide a sender which will fit into any available space however irregular this may be. In an operational weapon it is usually the space occupied by the warhead which is available for instrumentation during a trial firing. In the case of a particular equipment the space thus made available was cylindrical with a stout bulkhead at either end. This had to contain the telemetry sender and other instrumentation equipment. In order to mount this instrumentation it was necessary to use a cruciform structure in which the equipment fitted into the quarter segments.

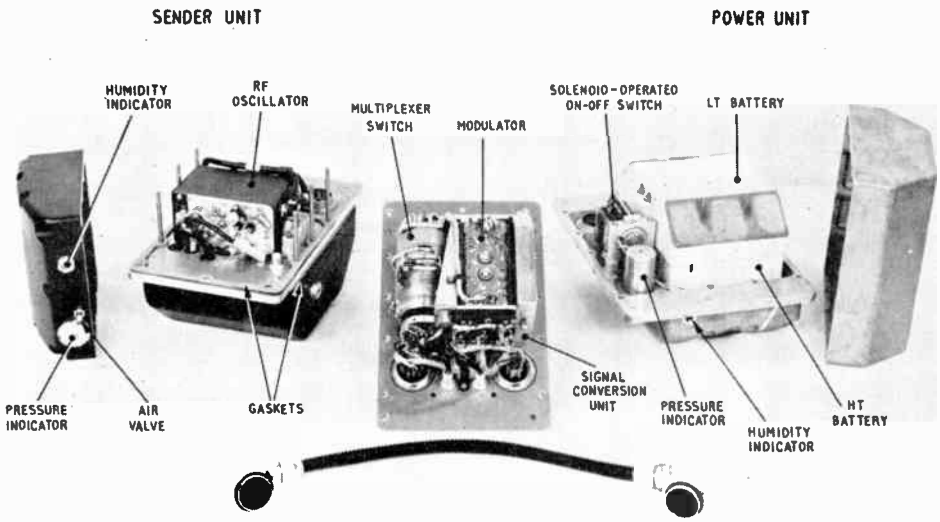


Fig. 7. Typical battery-powered sender.

Mercury primary cells are normally used for h.t. battery packs. They store more energy and have a very flat discharge characteristic (for a discharge current of 60 mA the potential falls from 1·15 V to 1·12 V after 5 hours) with a knee at their end-of-life voltage. The flat section of the curve is maintained even at high currents if the cell is kept warm (40° C). The minimum temperature without serious loss of watt-hour capacity is about 15° C, although at light currents the cell is usable well below -10° C.

The voltage depends on the load and temperature, but is of the order of 1·2 volts/cell and is suitable for assembly into h.t. batteries. The internal resistance is low and consequently a shock from an h.t. battery can be lethal.

The space allocated to the telemetry sender was part of two quadrants of the cruciform. Thus the outside configuration of the transmitting equipment was determined before the equipment design was started. Because of the available space the telemetry equipment had to be designed as two separate units; the sender and the power unit (Fig. 7). Both units were to be completely enclosed by metal covers to provide protection and screening. The necessary interconnections between them was to be contained in one cable form. The layout was further restricted by the need to make all other plugs and sockets readily accessible and to ensure that the cables, etc., when fitted to the units were within the outer limits specified.

The next consideration was to design the equipment to withstand the missile environment without any degradation in performance or reliability. The expected vibration spectrum was severe, ranging from vibration amplitudes of 10 g at 30 c/s to 70 g at 1 kc/s, being maintained at this relatively high level up to frequencies of the order of 7 kc/s. Steady acceleration of up to 50 g and decelerations of up to 50 g in 20 millisec were expected.

The sub-units, plus all the cable looms and plugs and sockets, had to be mounted in such a manner that they would survive the rigorous vibrational environment. The r.f. oscillator is the most vibration-sensitive component and must therefore have priority in mounting position. For this equipment the layout involved mounting all sub-units on either side of a flat, $\frac{1}{4}$ in. thick aluminium alloy plate. When all sub-units had been mounted the plate was vibration tested and the mechanical Q factor measured at critical points on the plate and sub-units. The initial test showed Q factors of up to 10. The main peaks occurred at about 350 and 700 c/s. These were considered to be the fundamental and second harmonic resonances of the unit.

A long series of experiments was undertaken to reduce these high Q values. Plates of tool steel and soft alloys were used. The vibration curves showed that a soft alloy casting of increased cross-section reduced the second harmonic appreciably. The addition of covers and gaskets reduced the Q to a reasonable level. As a final step, vibration tests were carried beyond the expected specification limit to the point of destruction.

It was decided that if both the transmitter and power supply units were made pressure-tight then the risk of coronary discharge (at great altitudes) occurring from high potential points would be removed. A pressure differential of approximately $2\frac{1}{2}$ lb/in^2 was used. Pressurized units with very low leakage rates were obtained by fitting gaskets between each cover and the mounting plate and by impregnating the covers which were alloy castings. To enable the state of the equipment to be readily checked during storage and before use, a pressure indicating device was fitted. A Schrader air-valve in each unit enables it to be pressurized by a foot-pump.

R.f. power in excess of 18 watts (peak) is produced when the transmitter is operating, some of which will appear as spurious radiation. Errors due to feedback from the oscillator to the sub-carrier modulator were eliminated by mounting modulator and oscillator on different sides of the base plate and by enclosing the modulator in a screening can. All power leads into the oscillator are decoupled. The second problem of spurious radiation is that caused by leakage through the gaskets. This is particularly dangerous from the safety aspect, because the r.f. radiation could initiate the missile break-up charge. In order to overcome this a gasket was developed in which aluminium gauze was embedded in natural rubber. When the gasket is clamped down the gauze just breaks the rubber surface and forms a complete electrical screen between the cover and the mounting plate. The use of these gaskets resulted in an external spurious radiation of less than 8 microwatts.

The ambient temperature in which the equipment is expected to operate could vary from $-15°$ C to $+70°$ C. At the upper limit it was difficult to keep the component temperature below a design target 100° C because of internal heat generation. The problems of heat dissipation were overcome by conventional methods—designing large heat sinks into the outer walls of the units and by spraying all outer castings matt black. The measured temperature rise inside the unit was found to be 12% to 15% less for matt black castings than for unpainted ones.

To reduce the moisture content of the air inside the equipment a desiccant container was fitted. Humidity indicators were mounted in a clearly visible position in the wall of both transmitter and power supply units. These contain a piece of chemically-sensitized paper, visible through a glass window, which changes colour with change of humidity.

5.2. *Electrical Design*

5.2.1. Sender

The major sub-units (oscillator, modulator, switch) are general purpose components but the signal conversion unit is unique for each type of sender. Its functions are threefold; firstly to adjust the level of the incoming signal to a suitable input for the modulator, secondly to adjust the impedance to a suitable value with respect to the modulator, thirdly to adjust the bandwidth of the incoming signal (using a simple low-pass filter). The unit is designed to give an error of not more than 1% under all conditions of environment. The inherent errors are due to two causes; the tolerance and drift in the values of resistors used, and the fact that the modulator input impedance and the source impedance of the signal transducer act as a potential divider when the multiplexer is switched to that transducer.

5.2.2. Power unit

For this particular application a separate battery power supply was the most satisfactory solution. Mercury-alkaline primary cells were used for the h.t. supply and nickel-cadmium cells for the l.t. supply.

The various h.t. and ancillary supplies are provided by 540 cells. The battery supplies the following potentials:

(1)	H.t.:	400 V, 200 V, 150 V
(2)	Bias:	5·3 V, −11·98 V

(3) Transducers: 6·6 V
(4) Calibration: 6·74 V, 2·7 V

These are stacked to conform to the unit configuration and each section is encapsulated. The sections are then mounted in a sealed polystyrene case. The operating temperature must be kept between 19° C and 50° C and to achieve this, heater mats are placed between banks of cells. A built-in thermal switch is used to keep the temperature at approximately 28° C. Heating supplies are provided from the launcher, so that heating can be continued until the missile is launched.

larger current is taken. These currents are measured on an ammeter—the deflection indicating the switch position.

Electrically the equipment has been designed to give stability of performance, in particular to be free from long-term frequency drifts of both sub-carrier and r.f. carrier. Calibration levels are transmitted to allow for slight changes in sub-carrier frequency. This enables the setting-up to be done in laboratory conditions, the units sealed, pressurized and installed in the missile, before it reaches the boosting and launching phase.

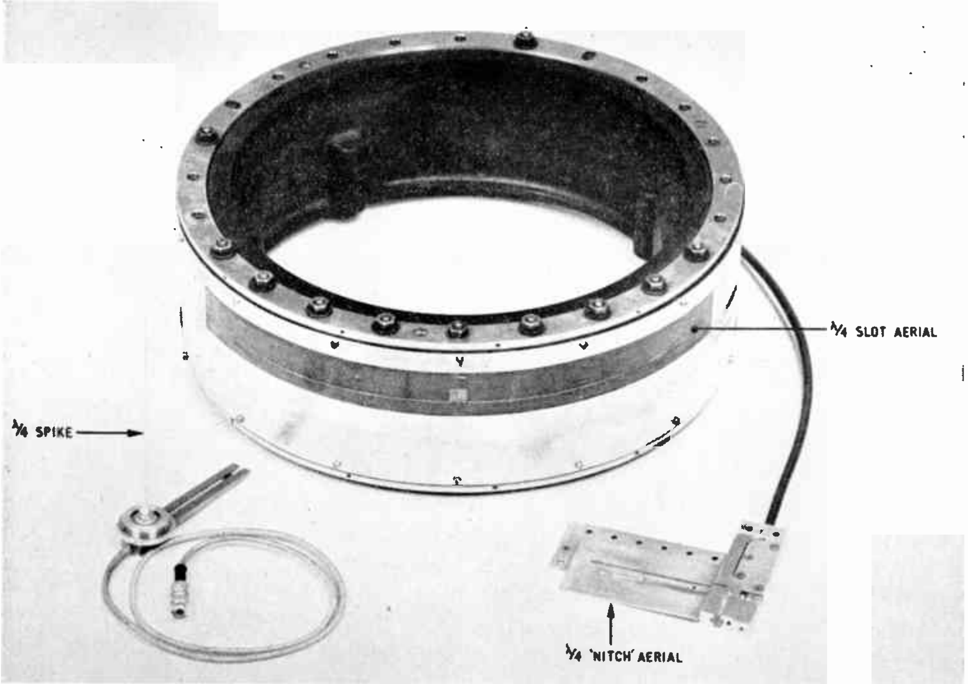


Fig. 8. Typical aerial systems.

The l.t. battery consists of five sintered-plate nickel-cadmium cells, connected in series to provide a nominal 6·5 V supply. Each cell is packed in a moulded polystyrene case and has two nickel-plated terminals.

In order to conserve the batteries it was necessary to be able to switch the sender power supplies on and off when the sender was mounted inside the missile. This was achieved by mounting a rotary wafer-switch with a single indexing solenoid (operating from a 24 V impulse). It was desirable to know when this electrically-actuated switch was "on" and when "off"; consequently a simple circuit was included in the power unit. With the switch in the "off" position the circuit takes a relatively small current from the impulse supply, in the "on" position a correspondingly

It is not necessary to re-align a transmitter after it has been thus set up and installed. The only limitation to the storage/ready-for-action time is imposed by the l.t. cells themselves, which have a reliable stand-by life in the charged condition of 7 days. After such a "setting-up to firing" delay the power unit would have to be removed from the missile and the l.t. batteries changed.

5.3. *Design of the Missile Aerial System*

The telemetry aerial installation for any missile must fulfil, as nearly as possible, three main requirements.†

† See, for instance, R. E. Beagles, "Telemetry aerials for high-speed test vehicles", *J. Brit.I.R.E.*, **18**, pp. 497–504, August 1958.

(*a*) It must be aerodynamically compatible with the missile.

(*b*) It must be correctly matched to the sender, so as to radiate as much of the sender output power as possible.

(*c*) It must radiate most of this power over a wide solid angle, producing a radiation pattern having as few nulls as possible within this angle. With such a radiation pattern changes of missile attitude will not produce excessive failing of the received signal.

Several types of aerial systems are suitable. The system used depends largely on missile configuration. In addition it is often necessary to achieve the last two requirements despite the change in missile configuration caused by boost separation from the missile. Aerial matching must be such as to give a voltage-standing-wave ratio of better than 0·6 over the telemetry band.

Two main types are briefly considered here: spike (or blade) aerials, and slot-type aerials (Fig. 8).

In most missile telemetry systems, the telemetry sender aerial system consists of two $\lambda/4$ elements placed diametrically opposite on the missile skin. Wherever possible the elements are placed at the position on the missile body most advantageous from the radiation pattern viewpoint, and are designed to minimize aerodynamic drag as far as possible. The aerial elements are fed as a balanced pair, i.e. the signals at the feed-point of each aerial are in anti-phase.

It is difficult to use conventional slot (or suppressed) aerials because they require a large driving cavity and thus can only be used on large missiles. Consequently the more usual slot aerial is a quarter-wave "notch" (or "nitch").

These are normally mounted in the trailing edge of the missile fins. Because of the position of the aerials at the rear end of the missile and the reflecting properties of the control surfaces, the radiation is predominantly in the rearward direction.

The "nitch" aerial and its matching stub slots are filled with polystyrene dielectric keyed into the rebate by means of countersunk holes on either side of each aperture. The impedance of each aerial is made nominally 100 ohms, then with an integral number of half wavelengths of cable between each nitch and the junction, the impedance at the junction becomes 50 ohms. Since the difference in length of these two cables is $\lambda/2$, the aerials are fed as a "balanced" pair.

As the shape of the missile and the positions of the aerials are fixed, the only variable parameters are the length of the aerial spike and the length of the short-circuited stub. Variations of the length of a spike aerial change both its resistance and reactance, but the change in reactance can be balanced out by adjusting the length of the stub. The resonant resistance of a quarter-wave aerial is between 30 and 35 ohms. If the length, L, is increased, the resistive component of impedance increases very readily—roughly as $(L/\lambda)^4$. The length is increased until this component is of the order of 50 ohms at the centre of the telemetry band. This value of resistance is required for optimum matching with the sender, but the increase in length also produces an inductive component. This is cancelled out either by adjusting the bushing capacitance at the base of the aerial or, if the base capacitance is already large, by means of a short-circuited stub.

The length of the aerial spike to be used is arrived at from a compromise between aerodynamic and electrical factors. Aerodynamically the length should be kept to a minimum, whereas to provide optimum impedance matching the spike needs to be longer than $\lambda/4$. However, it is not necessary to make the aerial present a perfect match. The major factors which will be influenced by a mismatch are:

(i) frequency pulling of the oscillator, and

(ii) power loss due to reflection.

Bearing these two factors in mind it is found that a v.s.w.r. of the order of 0·6 or better will still provide an adequate electrical performance.

6. Acknowledgments

The author wishes to thank the Ministry of Aviation and the Directors of E.M.I. Electronics Ltd. for permission to present this paper.

Manuscript first received 25th June 1959 and in revised form on 27th June 1960 (Paper No. 603).

# of current interest . . .

Commonwealth Technical Training Week

Preparations for Commonwealth Technical Training Week are now taking shape and the inaugural function which H.R.H. The Duke of Edinburgh is expected to attend, will take place at Guildhall, London, on 29th May. His Royal Highness will also attend a service at St. Paul's Cathedral on 1st June.

Other functions are being organized by the London County Council and in Edinburgh, Belfast and Cardiff supporting functions are being arranged. In Cardiff the formal opening of the Welsh College of Advanced Technology will take place.

Government Committee Inquiry into Higher Education

The Prime Minister announced in the House of Commons on 20th December that Professor Lord Robbins (Professor of Economics at the University of London) has been appointed chairman of a committee which is to inquire into higher education. The terms of reference of the committee are as follows:

"To review the pattern of full-time higher education in Great Britain and in the light of national needs and resources to advise Her Majesty's Government on what principles its long-term development should be based . . . whether there should be any changes in that pattern, whether any new types of institution are desirable and whether any modifications should be made in the present arrangements for planning and co-ordinating the development of the various types of institution."

The long-term development of Universities, Colleges of Advanced Technology, certain other colleges of further education and Teacher Training Colleges, will be within these terms of reference.

The Ministry of Education have announced that the appointment of this committee of inquiry will not hold up action on issues such as university expansion which are concurrently before the Government.

The Christopher Columbus Prize

As in previous years the President of the Institution was invited by the City of Genoa to submit a recommendation for the award of the Christopher Columbus Prize for 1960 on the outstanding development in space communications during the past quinquennium. The value of the Prize is five million lire (about £2750) and it is to be awarded to the International Astronautical Federation.

The citation of the award—based on a resolution of an *ad hoc* committee of the National Research Council of Italy—states that " The International Astronautical Federation has been the first and most important organization to promote the idea of astronautics in its scientific and technical aspects."

The British Radio Industry

Speaking at the Annual Banquet of the Radio Industry Council in November, Sir Thomas Spencer urged the Chancellor of the Exchequer to raise the level of personal income at which surtax applies. He believed that this action, which would result in negligible effect on the country's revenue, would give incentive and encouragement to engineers and scientists on whom the future of the country depends. Sir Thomas said that at present companies had to pay excessively high salaries to provide a reasonable standard of living for its senior scientific and executive staff.

Earlier in the evening statistics were quoted which showed that the increased production of the radio industry for 1960 was in the region of £400 million. The industry's exports of capital equipment amount to £87 million and a further £23 million was received from the export of entertainment equipment. It was stated that in June 1960 235 000 people were employed in the industry, an increase of 10% on the previous year.

Diversification at the College of Aeronautics

The Governors of the College of Aeronautics, Cranfield, have decided to change the titles of three of the College's teaching departments with effect from October 1961 as follows:

The Department of Aircraft Economics and Production is to be named the Department of Production and Industrial Administration.

The Department of Aircraft Electrical Engineering is to be named the Department of Electrical and Control Engineering.

The Department of Aircraft Materials is to be named the Department of Materials.

Since 1959 the College has been pursuing a policy of diversification of teaching activities and as the result a number of new advanced courses have been developed in subjects which, although relevant to aeronautics, are of equal interest and value to many other branches of engineering.

In addition to its normal function of providing specialist courses for aeronautical engineers as part of the two-year Diploma course, the Department of Electrical and Control Engineering now offers advanced courses in Industrial Control Engineering and Flight Control. The Department of Production and Industrial Administration holds advanced courses on, for instance, Operational Research and Ergonomics. The departments also provide specialist instruction in their subjects for students of the courses in Space Technology and Guided Missiles.

U.D.C. 621.398

455 Mc/s Telemetry Ground Equipment

Presented at a Symposium on Radio Telemetry, held in London on 25th March, 1959

By

F. F. THOMAS, B.SC.†

Summary: The 24-channel 455 Mc/s telemetry system has been the most widely used system in Great Britain for many years. This paper describes the principles common to all the present types of receiving and recording equipment in use on the ranges. Certain of the equipments are described in greater detail; these include receivers, a sixteen-tube slow-speed recorder and a histogram recorder with data processing facilities. The paper concludes with a description of a comprehensive test set for checking and calibrating telemetry senders.

1. Introduction

The purpose of the telemetry ground equipment is to receive, decode and record the information from the airborne sender. In this sender a 455 Mc/s carrier is amplitude modulated by a sub-carrier whose nominal frequency is 145 kc/s. The quantities to be measured vary this frequency over the band 130–160 kc/s.

Each measuring channel is connected to one segment of a motor-driven commutating switch which samples the channels sequentially, the wiper of this switch being connected to the transmitter modulator. The switch has 24 segments and samples at a rate of approximately 100 revolutions per second.

A synchronizing signal is transmitted by deviating the sub-carrier frequency to 186 kc/s on one of the channels. The transmission system is described in thc companion paper by W. M. Rae.[1]

The block diagram (Fig. 1) shows the connections between the various units of the ground equipment. The 455 Mc/s radio frequency signal from the sender is first passed through a receiver containing a demodulator for amplitude modulated signals: the output from this receiver consists of the frequency modulated sub-carrier signal. By means of limiters in the receiver this output signal is held at a constant amplitude of 3 volts peak-to-peak. The sub-carrier signal then passes to the discriminator and to the sync. separator. The discriminator demodulates the frequency modulated sub-carrier into a set of twenty-four d.c. levels which are fed to the main amplifiers of the recorders. This set of twenty-four d.c. levels is known as the histogram (Fig. 2).

The sync. separator separates and demodulates the sync. channel, producing a positive pulse output for each sync. channel input. These sync. pulses are used for generating strobes and markers for use with the recorders.

In the recorders cathode-ray tube displays are filmed by continuously moving film types of camera to give permanent records of the telemetry signal.

To prevent loss of records in the case of a failure in the ground equipment the whole equipment is usually duplicated.

2. Receivers

The r.f. signal from the sender is received on a helical aerial which tracks the missile's line of flight. This tracking may be manual or may be actuated by a servo-mechanism connected to a missile tracking system. This signal is then passed to the receiver via a 50-ohm cable.

2.1. *Receiver Type TD212*

This receiver incorporates a low-noise tuned r.f. amplifier, local oscillator, mixer and a 45 Mc/s i.f. amplifier with demodulator. The sub-carrier output from the demodulator is limited and filtered. A.g.c. voltage is derived from the sub-carrier and a signal level meter is operated from this circuit. A frequency deviation meter is operated by a discriminator supplied at 45 Mc/s from the i.f. amplifier (See block diagram, Fig. 3.).

The r.f. stage employs a planar disc-seal triode as a grounded-grid amplifier. Its associated resonant circuits are formed by capacitance-loaded transmission lines which are part of the mechanical assembly supporting the valve. The tuning of cathode circuit of the amplifier is preset, but the anode circuit is tuned by the panel control R.F. TUNE.

The anode circuit feeds a balanced mixer using a pair of germanium crystal diodes. The local oscillator is another disc-seal triode and oscillates at a frequency 45 Mc/s above the signal frequency. It is tuned by the panel control OSC. TUNE.

The i.f. unit has six stagger-tuned amplifier stages plus a diode demodulator and an output cathode

† J. Langham Thompson Ltd., Stanmore, Middlesex.

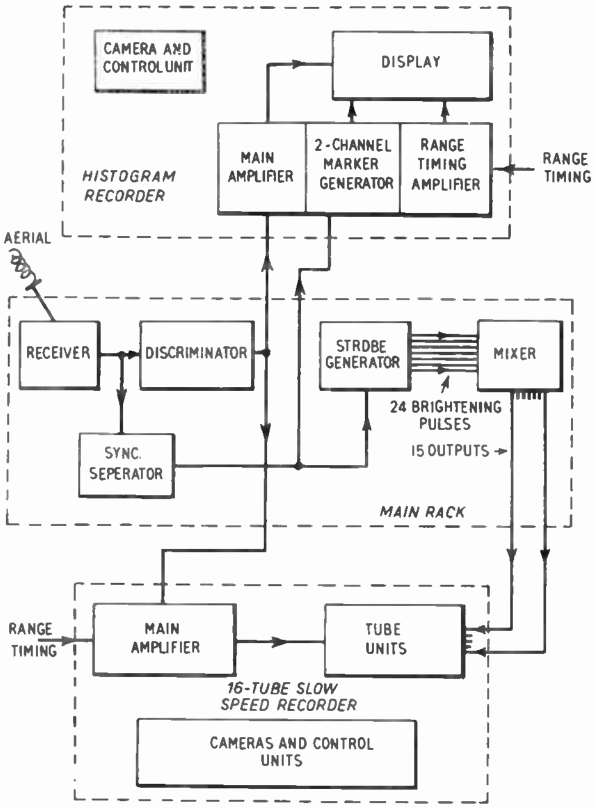


Fig. 1. Block diagram of 455 Mc/s telemetry ground equipment.

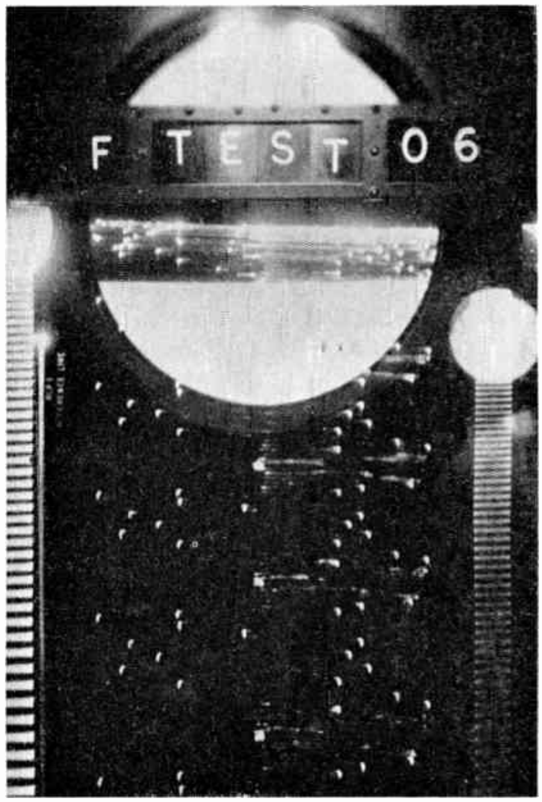


Fig. 2. Photograph of histogram.

follower. A 45 Mc/s signal is taken from the last i.f. amplifier valve to feed the discriminator which operates the frequency deviation meter.

The sub-carrier signal from the i.f. amplifier output cathode follower is amplified and then limited to 3V peak-to-peak. This limited signal is fed to the sync. output socket via a cathode follower and also to the input of a band-pass filter. This filter has a pass-band of 125 kc/s to 165 kc/s and further improves the signal/noise ratio in the information band.

The a.g.c. circuit is driven by the sub-carrier signal at the output of the i.f. amplifier. A peak-to-peak rectifier circuit is employed to eliminate any variations due to changes of mark/space ratio of the sub-carrier signal. The a.g.c. output socket provides a low impedance d.c. output which is used to drive a signal strength recorder.

2.2. *Receiver Type TD*240

This receiver is used where the sensitivity of the receiver type TD212 is not required. The r.f. signal is injected directly into a flat tuned mixer so that the only tuning control necessary is the local oscillator frequency control. The i.f. and limiter stages are identical with those in receiver type TD212 but the discriminator circuit and frequency deviation meter are not fitted. The receiver type TD240 and receiver type TD212 are mechanically interchangeable and their external power requirements are identical.

Table 1 gives a comparison of the two receivers.

3. Discriminator and Sync. Separator

3.1 *Discriminator*

The discriminator converts the 100 kc/s–200 kc/s sub-carrier signal into a d.c. output signal. Several types have been used over the past ten years, but they have all been variations of the "pulse integration" type. In this type of discriminator the incoming sub-carrier is amplified and squared and a constant-area pulse then generated once per cycle of the incoming waveform. These pulses are integrated to give a mean d.c. level proportional to the input frequency.

The most significant advantage of this type of discriminator is that it is inherently linear over the whole of its frequency range. Tuned circuit type discriminators have a small working bandwidth and the linearity obtainable is largely a function of the Q of the components and the skill with which they are lined up.

In the different types of discriminator that have been used the constant area pulses have been generated by flip-flop pulse generators, delay lines or constant charge and discharge capacitor circuits. There is little to choose between the results obtained from properly designed discriminators using delay line or constant charge and discharge capacitor circuits, but the flip-flop pulse generator is not so satisfactory particularly because of its poorer performance in the presence of noise.

signal, since any variation of this trigger point gives the same effect as a frequency or phase modulation of the input signal and hence appears as unwanted output. For this reason the amplifier and limiter selects a narrow slice of the waveform about the zero point when a.c. coupled and amplifies it into the triggering square wave. Circuits which tend to restore the d.c. level of the input waveform must be avoided as these can cause a variation in the trigger point when amplitude modulation is present.

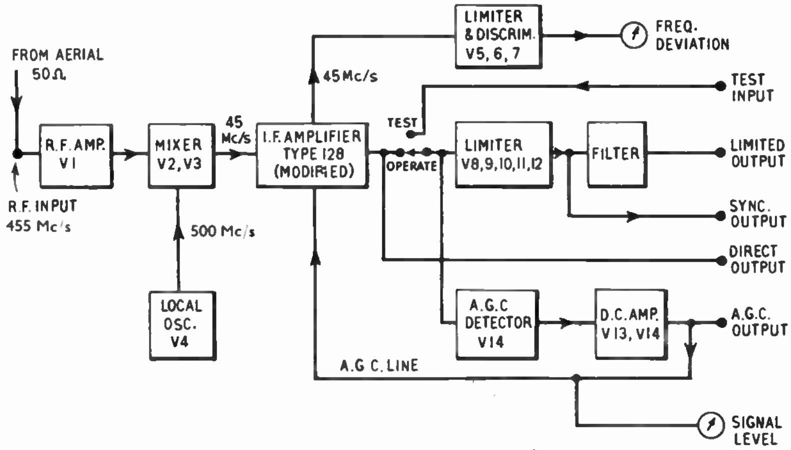


Fig. 3. Block diagram of telemetry receiver TD212.

The design of discriminators consists of three basic stages: the amplifier and limiter, the pulse generator and the integrating filter and d.c. output amplifier.

It is important that the pulse generator should always be triggered from the same phase point of the input

The foregoing considerations led to the introduction of the limiter circuit shown in Fig. 4. In the quiescent state both halves of the double triode are passing approximately equal anode currents. During the positive half-cycle of the input waveform the first triode acts as a cathode follower and the second

Table 1

	TD212	TD240
Nominal Tuning Range	30 Mc/s range in 400–500 Mc/s band	410–500 Mc/s
Bandwidth	2·5 Mc/s–3 Mc/s	3·5 Mc/s
Average Noise Factor (50 ohm source)	8·5 dB	11 dB (including 3 dB image noise)
R.F. Input for 1 : 1 Signal/Noise Ratio	Approx. 4 μV	Approx. 10 μV
Overall Gain	Approx. 100 dB	80 dB minimum
Limited Output Level	3 V peak-to-peak	3 V peak-to-peak
Direct Output Level (before Limiters)	200 mV–2 V peak-to-peak	200 mV–2 V peak-to-peak
A.G.C. Characteristic	10 dB output rise for 40 dB input rise	10 dB output rise for 40 dB input rise

In the Type TD212 the frequency deviation meter is linear over the range ± 2 Mc/s. The meter error is less than 0·25 Mc/s providing the signal/noise ratio is greater than 20 dB.

triode cuts off when the cathode rises to a voltage greater than the grid base of the valve. During the negative half-cycle of the input waveform the first triode is cut off and the second triode passes an anode current determined only by the values of the anode load and cathode resistors. A number of limiter stages are used in series. The first stage is designed to handle a wide range of input voltages without running into grid current and the subsequent stages are designed for maximum gain consistent with the ability to handle the known limited output of the first stage. In most pulse generators the rate of rise and shape of the trigger pulse has some effect on the area of the output pulse. In order that the pulse generator shall always have the same shape trigger the limiter circuits must have sufficient gain to ensure that the rise time of the output square wave is dependent only upon the time-constant of the limiter output circuit and independent of the input waveform rise-time.

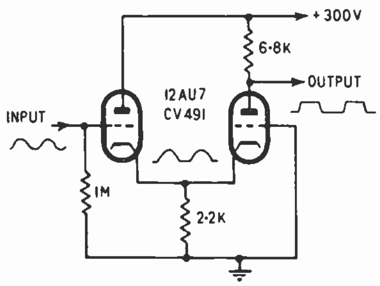


Fig. 4. Basic limiter circuit.

The circuit of Fig. 4 is simple and does not suffer from d.c. restoration effects but due to the slightly different operating conditions and grid bases of the two halves of the valve the output square wave is not of exactly 1 : 1 ratio. This can be corrected by providing a variable bias adjustment to the grid of the second triode. A further refinement is to apply d.c. negative feedback to each half to stabilize the operating points. A full description of such a circuit is to be found in reference 2.

The pulse generator must have good stability of pulse area and the circuit must completely recover to its quiescent state in a time less than the time required for one complete cycle of the highest input frequency. The presence of a "memory" time-constant in this circuit will result in the output level of a particular channel varying with the frequency of the previous channel. This is known as "switching error" and is defined by

$$\text{Switching error } \% = \frac{(f_{\text{apparent}} - f_{\text{single channel}})}{30 \text{ kc/s}} \times 100$$

where $f_{\text{single channel}}$ is the frequency indicated by the discriminator with a continuous sine wave input, and f_{apparent} is the frequency indicated by the discriminator with the same input frequency on one of its twenty-four channels and any frequency in the range of 130 kc/s–160 kc/s on the previous channel. Some early discriminators had switching errors of up to 3%.

Any variations of pulse area caused by supply voltage variations directly affect the output sensitivity, and for this reason the h.t. and heater supplies are often stabilized.

The integrating filter must remove all sub-carrier frequency components of the pulses and yet have a wide enough pass-band to reproduce the histogram faithfully. If the filter has too low a cut-off frequency the histogram channels will rise too slowly and if the filter has too sharp a cut-off characteristic ringing will appear on the channel top. With a correct design the channel will rise to its final value in less than one half of a channel width and then remain flat without ringing until the end of the channel.

The d.c. amplifier is designed for low drift and its gain and d.c. reference point are made adjustable in order that the output can be set to zero at 145 kc/s input with an output/input characteristic of 0·1 volt/kc/s.

The specification of the present discriminators over the range 130 kc/s–160 kc/s is:

Linearity (Deviation of any point from best straight line) <0·3%.

Switching error <1%.
Better than 50% flat channel top.
Sensitivity 0·1 volt/kc/s.

3.2. *Sync. Separator*

The sync. channel is separated by passing the sub-carrier signal through a tuned amplifier-limiter. The tuned circuit frequency (180 kc/s or 186 kc/s) is selected by a switch according to the channel in use. The sync. frequency output of this circuit is then demodulated and the demodulated pulse used to switch on a normally cut-off valve. The Q of the tuned circuit and the valve bias are adjusted so that the sync. separator only responds to sync. frequencies within $\pm 2\frac{1}{2}$ kc/s of the nominal sync. frequency. This is necessary to prevent false triggering when more than one sender and sync. frequency are in use at the same time.

4. Histogram Recorder

The output from the discriminator is passed directly to the histogram recorder. This recorder serves two purposes. The first is to provide a record of the received signal after the minimum amount of processing. This enables the discriminator output to be examined in detail and prevents loss of information should later ground circuits fail or the signal be noisy. The second

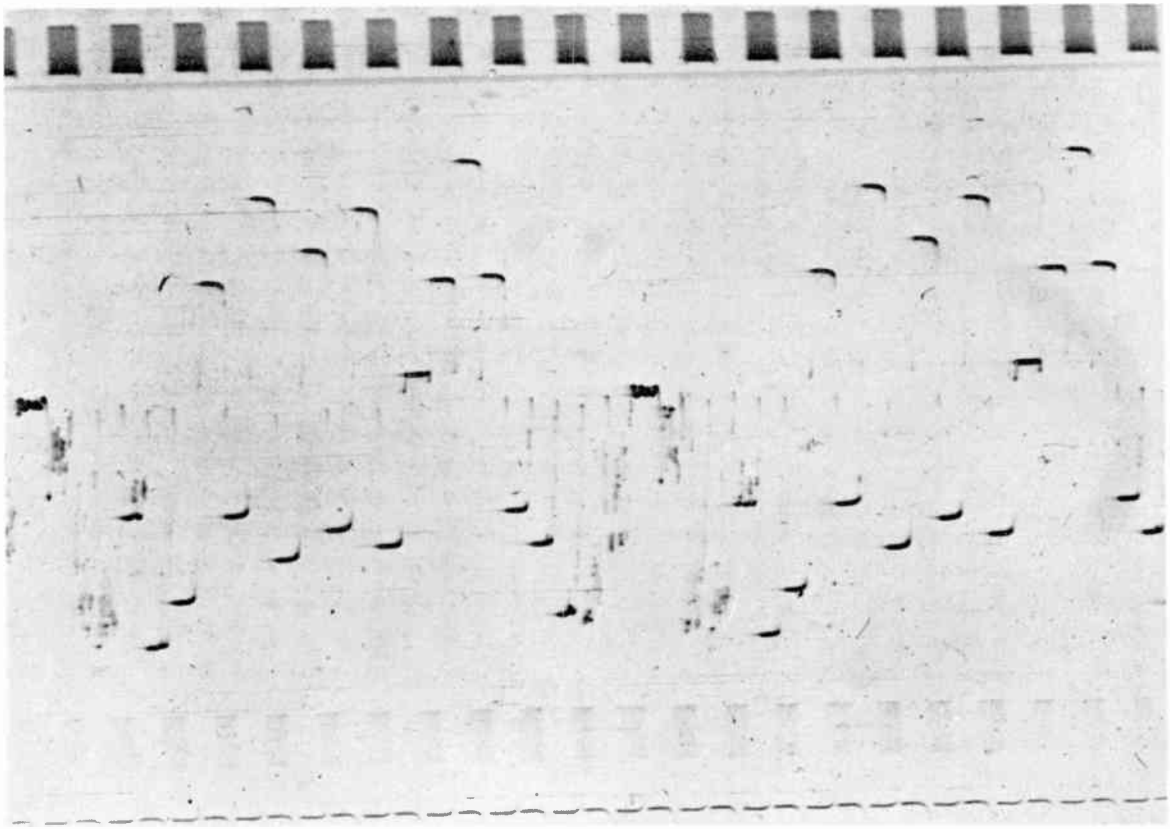


Fig. 5. Photograph of histogram film.

purpose is to provide a detailed and accurate record that can be decoded by automatic data processing equipment.

4.1. *The Histogram Recorder Traces*

The record consists of five traces on 35 mm film (Fig. 5). The large central trace is the histogram. To produce this trace the output from the discriminator is amplified until the limits of the information channels (130 kc/s–160 kc/s) occupy the whole working diameter of the recording tube. The sync. channel deflects off the face of the tube and so the top of this channel is not visible. This d.c. amplifier has been designed so that there is no interaction between the shift and gain controls.

A reference line is put on the film to form a base from which to take measurements of the histogram information. This line is formed by concentrating the light from a pilot lamp bulb through a perspex cone to form a bright pinpoint of light on the display front panel.

The uppermost trace is the millisecond marker which consists of one black and one clear mark per millisecond. In the data processing equipment a photoelectric cell senses these marks, and, working in conjunction with a counter, enables the equipment to start data processing at any point on the film to within a millisecond of the original timing. This marker is obtained by generating pulses one half millisecond wide from the one millisecond interval range timing pulses. These $\frac{1}{2}$ millisecond pulses are then applied to the grid of a cathode-ray tube as bright-up signals. The width of the markers is obtained by applying a 150 kc/s spot wobble signal to the vertical deflection plates of the cathode-ray tube.

The trace immediately below the histogram is the two-channel marker. This trace consists of one black and one clear mark per two channels of the histogram signal, i.e. each black or clear mark is one channel in length. The width of these markers is obtained by the same spot wobble signal as the millisecond markers. The method of generating the two channel marker pulses and the use of this marker in the data processing of the film is described in Section 4.2.

The bottom trace on the film is the normal range timing signal. This will vary in content from range to range but a typical signal consists of a "ruler" display of milliseconds, $\frac{1}{100}$th seconds, $\frac{1}{10}$th seconds, and 1 second interval pulses. This timing trace is used when visually examining the film.

4.2. *The Two-channel Marker and Data Processing*

4.2.1. An outline of the data processing system

The automatic film reader requires a film that has clear lines on a black background and that has each "frame" from sync. pulse to sync. pulse of a specific length. To achieve this the original record is first transferred to another film. The film is printed by moving the original film and the positive stock it is to be printed on across a slit of light. The original film is moved at constant speed but the positive stock is moved by a servo-controlled motor at such a speed as to expand or contract the original "frame" length into the correct length for the film reader.

A photo-electric cell senses the two-channel markers on the original film and compares the signal obtained with another signal obtained from the sprocket holes of the positive stock. In order to obtain two signals of the same frequency the sprocket hole signal is multiplied by six before comparison with the two-channel marker signal. Any frequency error between these signals is fed into the servo system controlling the speed of the positive stock. As lag in the servo cannot be avoided the delay between sensing and correcting is padded out to a known time. On the histogram recorder the two-channel marker tube is displaced to one side of all the other recording tubes. Thus the two-channel marker is put on the film a fixed distance ahead of the channel it refers to. At the printing speed this distance corresponds to the servo lag plus the delay time.

If the two-channel markers cease to appear on the original film the printer stops and operates an alarm. The two-channel markers do not appear on the printed film, but if the markers have ceased a distinguishing mark is put on the film which stops the reader on reaching that point.

In the reader the film is placed over the face of a cathode-ray tube and the sync. pulse aligned with a spot on the tube. After this the film is moved on frame by frame. A servo system senses whether the sync. pulse is aligned with the spot and if necessary makes a small adjustment to each frame position.

As each frame is placed in position a number of spots traverse the film. Photocells are placed above the film and as each spot traverses the reference line and the top of a telemetry channel an output pulse is obtained from the photocells. The spacing between the two pulses of a pair is a measurement of the value of the channel being traversed.

The results of these measurements are put on punched cards which are suitable for feeding a computer or automatic plotting table.

4.2.2. The two-channel marker

A block diagram of the two-channel marker is shown in Fig. 6. The actual marker generator is a symmetrical multivibrator running at twelve times the switch motor frequency (i.e. 1200 c/s for a 100 rev/sec switch motor). As the only information available as to change of speed of the switch motor is derived from

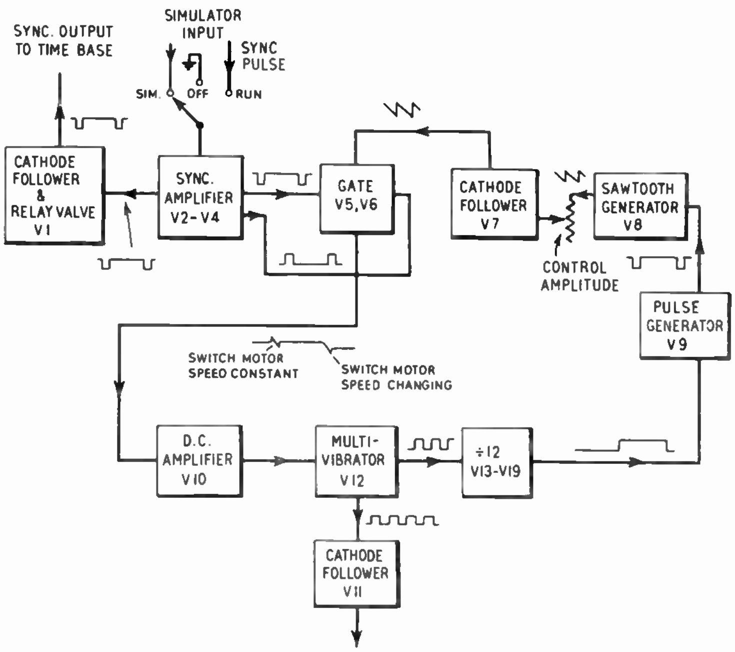


Fig. 6. Block diagram of the 2-channel marker.

the arrival of the sync. pulse once per revolution, to keep the markers in step with the channels their frequency is divided by twelve and compared with the sync. pulse.

As shown on the block diagram the multivibrator output is divided by twelve and a narrow pulse generated from the divider output. This pulse is used to trigger a saw-tooth generator whose output is applied to a gate circuit. This output is symmetrical about earth potential. The gate is only opened during the sync. pulse period, and whilst it is open the saw-tooth voltage is applied to the input of a d.c. amplifier. This amplifier has a memory circuit in its grid so that when the gate closes the grid voltage remains constant until the gate reopens.

If the switch motor speed in the sender is constant the gate will always open at the same point on the saw-tooth and a constant level will be fed to the amplifier. If a change in motor speed occurs, the gate will open at a different saw-tooth voltage level and a step correction voltage will be fed to the d.c. amplifier. After the gate has closed the amplifier "remembers" this step until the next sync. pulse opens the gate. The output of the d.c. amplifier controls the multivibrator frequency thus completing the servo loop.

The incoming sync. pulses are also applied to a relay circuit which controls the h.t. supply to the multivibrator. If three consecutive sync. pulses are absent the relay operates and switches off the two-channel marker multivibrator. As soon as the sync. pulses reappear the relay is switched off and the two-channel marker operates once more.

5. Strobe and Mixer Circuits

Although all the information transmitted by the sender is present on the histogram recorder record it is not in a form that can be readily assimilated by visual examination. For this reason a strobe circuit is employed that separates each information channel. The consecutive levels of each channel for each revolution of the switch motor can then be displayed as a separate trace on a film.

5.1. *Strobe Circuits*

The sync. pulse from the sync. separator is used to trigger a saw-tooth generator, the output of which is connected to twenty-four strobe circuits. As the saw-tooth voltage passes the trigger voltage of each strobe circuit an output pulse is produced from the strobe. The twenty-four trigger voltage levels are adjusted so that as each strobe triggers during the saw-tooth run-down its output pulse corresponds in time with the flat top of one of the telemetry channels.

In order to prevent the strobes from moving along the channel tops if the switch motor speed varies a feedback system is employed. As the time-base runs at a constant rate any variation in the time of arrival of the sync. pulse alters the time-base amplitude. A reference voltage derived from the time-base amplitude is fed to the strobes and their trigger points are automatically compensated for amplitude changes.

Each strobe has to be individually adjusted for position to compensate for positional errors in the switch motors. Contemporary switch motors however, are considerably better in this respect and in the future it might be possible to provide an automatic strobe system without individual adjustment.

5.2. *Mixer Circuits*

In practice the output from more than one channel may be needed on some or all displays, and comparison with calibration channels and with channels giving information on the operation of associated equipment is a common requirement. To meet these needs the twenty-four strobes are fed into mixer circuits which enable any number of the twenty-four input strobes to be mixed on to any one of fifteen output lines.

In the earlier equipments the selection was by means of a bank of switches with twenty-four switches in each of fifteen rows. This system has been replaced in more modern equipments by a punched plate system. In this system spring-loaded contacts press against an insulated plate. A code of holes is cut in the plate corresponding to the channels required and the appropriate contacts pass through the holes to connect with another set of contacts on the other side of the plate. This system is described more fully in a companion paper by F. G. Diver.[3]

6. The 16-Tube Slow-speed Recorder

This recorder is shown in Fig. 7. The sixteen tubes are grouped in four groups of four tubes and each group is photographed by a type G.W.5 camera using 5½ in. film. In Fig. 7 three of the cameras have been removed in order to give a clearer view of the racks. The film speed of these cameras is adjustable from 1 in./sec to 16 in./sec in five steps, and the cassettes hold 200 ft of film. A magnifying periscope may be inserted through the top plate of the camera for critical focusing and a footage indicator is provided to show how much film has been used.

Fifteen of the cathode-ray tubes are connected to the fifteen mixer outputs of the main equipment. The strobe pulses from the mixers are used to brighten the tubes which are normally cut off. The discriminator output signal is amplified by a d.c. amplifier and fed to all fifteen vertical deflection circuits in parallel. Thus the beams of all the tubes follow the histogram signal and the strobe pulses produce a bright spot on the screen at a time when the deflection corresponds to the flat top of the appropriate channel.

The sixteenth tube is a spare and has its own vertical deflection amplifier. It can be used for recording supplementary information, e.g. receiver signal strength.

Timing signals at $\frac{1}{10}$th second intervals are recorded by deflecting all the cathode-ray tube beams to a point outside the normal range of information signals and simultaneously applying a brightening pulse to all tubes. Further timing signals at $\frac{1}{5}$th and 1 second intervals are recorded by momentarily flashing neon tubes situated between the recording cathode-ray tubes.

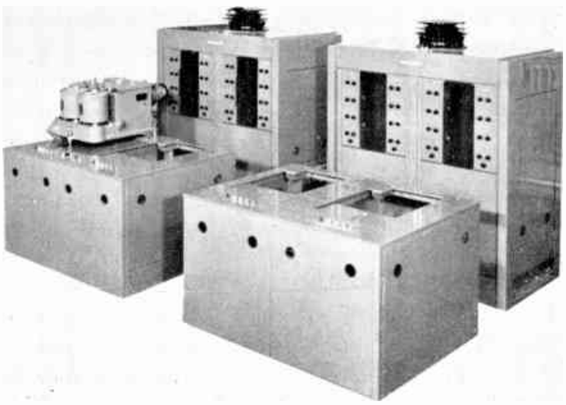


Fig. 7. The slow-speed recorder.

A pilot lamp is fitted above and below each group of four tubes and the light from each of these lamps is concentrated to a point source by means of a perspex cone. These lamps produce two reference lines on each film record which are used as a base for measurements on the recorded signals.

7. Camera Controls

All the recorders contain a camera control unit to programme the switching on and braking of the camera motor. The same design is used for both 35 mm and $5\frac{1}{2}$-inch film cameras and is used with film speeds ranging from 1 in./sec to 96 in./sec.

Local or remote control may be used but normally remote control is used on the ranges. The starting instruction is received from the range programme circuits. This is sent out two seconds before zero time to enable the cameras to run up to speed before the instant of firing.

On receipt of the start instruction the camera shutter is opened and the titling lamps switched on. After a delay of approximately 40 milliseconds the titling lamps are extinguished and the camera motor started. The camera then runs until the start instruction is terminated. Immediately this occurs the shutter is closed and the motor terminals are disconnected from the mains input and connected across the output terminals of a selenium rectifier. This rectifier absorbs any switching surges produced in the motor. After approximately 40 milliseconds the input to the rectifier is connected to the 30-volt winding of a mains transformer. This applies 24V d.c. to the motor winding and the motor brakes sharply. After 2–3 seconds the 24V d.c. is removed and the control unit is returned to the quiescent state.

In setting up the recorders a time-base is used on the horizontal deflection circuits of the cathode-ray tubes. To prevent the record being ruined should this time-base be inadvertently left running whilst recording, the relay that initiates the camera start also switches off the time-base via interlocking circuits. Similarly in some recorders the e.h.t. for the cathode-ray tubes is normally switched off to prevent the spots burning the screen phosphor. In these recorders the camera start relay also switches on the e.h.t. circuits.

8. Telemetry Test Set TD240

This equipment is used for testing and calibrating 455 Mc/s telemetry senders. The sender output at either r.f. or sub-carrier frequency is detected and displayed against various calibrating signals. Figure 8 is the block diagram of the test set which is shown in Fig. 9.

The receiver is a normal Type TD240 or TD212 as described in Section 2. The output from the receiver is passed through a gate circuit to the discriminator and thence to the vertical deflection circuits of the display.

8.1. *Calibration Circuits*

The calibration signals are derived from six crystal oscillators and a precision variable frequency oscillator. The crystal oscillators are at 110 kc/s, 130 kc/s, 145 kc/s, 160 kc/s, 180 kc/s and 186 kc/s. In order to display these signals visually with the sub-carrier output of the receiver, they are passed to the discriminator via gate circuits.

There are four of these gate circuits and by means of push buttons on the front panel of the gate and discriminator the following modes of operation are available:

(*a*) Gate 1 continuously open.

(*b*) Gate 2 (sub-carrier signal) continuously open.

(*c*) Gate 3 (130 kc/s crystal signal) continuously open.

(*d*) Gate 4 (160 kc/s crystal signal) continuously open.

(*e*) Gates 1 and 2 open on alternate time-base strokes.

(*f*) Gates 1, 2, 3 and 4 each open on sequential time-base strokes. The input to gate 1 is controlled by a switch on the oscillator unit and may be either 110 kc/s, 145 kc/s, 180 kc/s, 186 kc/s or the variable frequency oscillator.

Thus if the sixth button (*f*) is pressed the display shows sequentially gate 1, sub-carrier, 130 kc/s and 160 kc/s. To the eye this display appears as a histogram with calibration lines drawn across it. The frequency of any channel may be found by aligning the variable frequency oscillator line with the top of the channel and reading off the oscillator frequency.

Each gate circuit has an associated neon indicator lamp and the illumination of the lamps indicates which gate is open at any moment.

8.2. *Time-base Circuits*

The display time-base may be either free running, triggered from the sync. separator or delay triggered from the sync. separator. The purpose of the delay triggered operation is to allow the histogram sync. pulse to appear near the centre of the screen should it be desired to examine it.

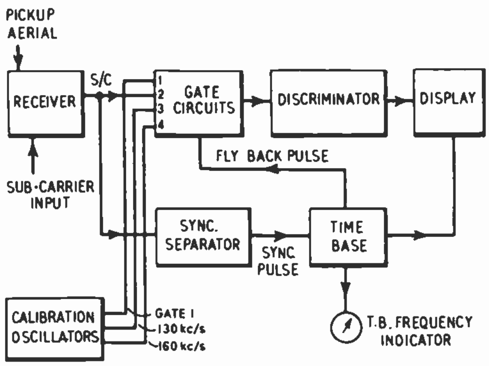


Fig. 8. Block diagram of 455 Mc/s telemetry test set.

With all modes of operation an electronic frequency meter measures the repetition rate of the fly-back pulses. When the time-base is in the trigger or delayed trigger modes this meter gives a direct reading of the switch motor speed in the sender.

In both the trigger and delay trigger modes of operation d.c. feedback is applied from the frequency meter circuit to the time-base rate, and delay-time circuits. This feedback compensates for any change of switch motor speed and keeps the time-base length on the screen and the position of the delayed sync. pulse constant.

The time-base fly-back pulse is also used to drive a pulse generator which triggers the gate circuits preceding the discriminator.

Fig. 9. The 455 Mc/s telemetry test set type TD240.

8.3. *Deflection Circuits*

The vertical and horizontal deflection circuits each consists of a d.c. amplifier with coarse and fine gain controls. The coarse gain controls give gain steps of $\times 1$, $\times 3$ and $\times 9$. With these controls in the $\times 1$ position the fine gain controls enable the histogram size to be adjusted so that the twenty-four channels horizontally and the information and sync. levels vertically just fill the screen. The horizontal and vertical shift controls and the gain controls are so arranged that whichever part of the waveform is shifted to the centre of the screen remains in the centre for all positions of the gain controls.

9. Acknowledgments

The author wishes to thank the Ministry of Aviation and the Directors of J. Langham Thompson Ltd. for permission to publish this paper.

10. References

1. W. M. Rae, "Engineering aspects of missile telemetry equipment", *J. Brit.I.R.E.*, **21**, pp. 57–67, January 1961.
2. Y. P. Yu, "Coincident slicer measures phase directly", *Electronics*, **31**, No. 37, pp. 99–101, 12th September 1958.
3. F. G. Diver, "Transportable ground receiving and recording equipment for 24-channel telemetry", *J. Brit.I.R.E.*, **20**, pp. 457–64, June 1960.

Manuscript first received 20th July 1959 and in final form on 27th September 1960 (Paper No. 604).

APPLICANTS FOR ELECTION AND TRANSFER

As a result of its January meeting the Membership Committee recommended to the Council the following elections and transfers.

In accordance with a resolution of Council, and in the absence of any objections, the election and transfer of the candidates to the class indicated will be confirmed fourteen days after the date of circulation of this list. Any objections or communications concerning these elections should be addressed to the General Secretary for submission to the Council.

Transfer from Associate Member to Member

BURBIDGE, Stanley Richard. *Brighton, Sussex.*
TURNER, Lewis Edgar. *Hatfield, Herts.*

Direct Election to Associate Member

BALL, John Geoffrey. *Surbiton, Surrey.*
BEER, Kenneth William. *London, S.E.26.*
GARLICK, Norman Laurence, M.Sc. *Gloucester.*
HILLS, Raymond Clement, B.Sc.(Eng.). *Welwyn, Hertfordshire.*
HOPKINS, Ronald James. *Chesham, Buckinghamshire.*
LANSDELL, Frank. *Stevenage, Hertfordshire.*
LOCKYER, Daniel Ernest Crittal. *Shanklin, Isle of Wight.*
SKELTON, Ronald George, B.Sc. *Bracknell, Berkshire.*
SLY, Maurice Edward. *Twickenham, Middlesex.*
WILCOX, Maj. Ernest Richard James, R.E.M.E. *London, S.W.1.*

Transfer from Graduate to Associate Member

CRAPPER, David Hugh. *Lovedean, Hampshire.*
DUNLOP, Alastair Donald. *Sandhurst, Berkshire.*
ERRINGTON, George Shacklock, B.Sc. *Christchurch, New Zealand.*
FREEMAN, Kenneth George, B.Sc. *Redhill, Surrey.*
HORN, Peter Jack. *Hitchin, Herts.*
KEY, Alfred John, B.Sc., *Dorchester, Dorset.*
LAM, Yat Wah. *Edgware, Middlesex.*
MILLER, David Peter. *Virginia Water, Surrey.*
OLSEN, George Henry, B.Sc. *Newcastle-on-Tyne.*
PASCOE, Thomas Albert. *Adelaide, South Australia.*
RICKETTS, Peter William. *Maidenhead, Berks.*
RUBIN, Moshe. *Haifa, Israel.*
THOMPSON, Philip Martin. *Romsey, Hants.*
TOWELL, Roger Percival. *Chalfont St. Giles, Buckinghamshire.*
WILLIAMS, Peter Brundall. *Pinner, Middlesex.*

Transfer from Student to Associate Member

BAMFORD, Thomas Arthur. *London, N.22.*
TYLER, Ronald Arthur. *Clacton-on-Sea, Essex.*

Direct Election to Associate

ERRIDGE, Frederick Fulford. *St. Laurent, Canada.*
FANEYE, Theophilus Alabi. *Lagos, Nigeria.*
GLASSCOCK, Eric Marcel. *Jamaica, West Indies.*
HALL, John Emmerson. *Nottingham.*
HEMMER, Hugh Donald. *Stoke-on-Trent, Staffordshire.*
JAMES, John Augustine. *Dublin, Eire.*
LEIGHTON, John. *Chalfont St. Peter, Buckinghamshire.*
THORPE, Arthur Sydney. *Hatfield, Hertfordshire.*

Transfer from Student to Associate

CROLE-REES, David George, M.A. *Geneva, Switzerland.*

Direct Election to Graduate

BADGER, Terence James, B.Sc.(Eng.). *London, N.W.4.*
BELL, Flt.-Lt. Angus Macdonald, B.Sc., R.A.F. *Singapore.*
COOK, David Alan. *London, N.4.*
CREE, Ronald George, B.Sc. *Kingston-on-Thames, Surrey.*
DYKE, John Roger, B.Sc. *Wembley, Middlesex.*
GARRAD, Capt. Roger Allan. *Malvern Wells, Worcestershire.*
GRIFFITHS, John, B.Sc. *Merthyr Tydfil, Glamorgan.*
HALL, Leo Frederick. *Nairobi, Kenya.*
HANSFORD, Richard Leslie. *Plymouth, Devon.*
HEASON, Brian. *Stapleford, Nottinghamshire.*
HUGHES, Raymond. *Hounslow, Middlesex.*
JOHNSON, James Samuel. *Colchester, Essex.*
KEATS, Albert Brian. *Dorchester, Dorset.*
LUSKOW, Alfred Allan. *St. Albans, Hertfordshire.*
MANNING, Robert Christopher. *Hayes, Middlesex.*
MAYHEW, Albert Francis. *Chelmsford, Essex.*
MILES, Denis Evelyn, B.Sc.(Eng.). *Hythe, Kent.*
MONTGOMERY, John. *Thatcham, Berkshire.*
MORGAN, Geoffrey Edward. *Sutton, Surrey.*
NASH, Francis Peter. *Dorchester, Dorset.*
SAPSFORD, George. *Basingstoke, Hampshire.*
STOCKER, Rex Thomas. *Weston-super-Mare, Somerset.*
TAGG, Frederick George. *Hemel Hempstead, Hertfordshire.*
TAYLOR, Jack. *Chelmsford, Essex.*
WILLIAMS, Peter Albert. *Bolton, Lancashire.*
WOLFE, Michael Anthony, B.Sc. *Warsash, Hampshire.*
WOOD, Ivor Thomas, *Stockport, Lancashire.*

Transfer from Student to Graduate

BAIRSTOW, Jeffrey Noel. *Sheffield, Yorkshire.*
BALDWIN, John Richard. *Knaresborough, Yorkshire.*
CATER, Michael William. *Walsall, Staffs.*
DU BARRY, James Joseph. *Dun Loaghaire, Eire.*
FOOKES, Reginald Arthur. *Caringbah, Australia.*
GOLDSMITH, Geoffrey Grant. *Beckenham, Kent.*
HILL, John William. *Newport, Monmouthshire.*
ISAAC, Ponnazhath Varghese, B.Sc. *Shertally, India.*
JAYATUNGE, Nihal Gamini Perera. *Colombo, Ceylon.*
LOYNES, David Howard. *Liverpool.*
PENNICOTT, Lloyd Hale. *London, W.3.*
RAE, Alexander Watson. *Glasgow.*

STUDENTSHIP REGISTRATIONS

The following students were registered at the October and November meetings of the Committee. The names of a further 57 students registered at the November meeting together with 55 students registered at the January meeting will be published later.

McPHEE, Plt. Off. K. J., R.A.F. *Notts.*
MANIKKAVACHAKAN, C., B.Sc. *Enfield.*
MEADOWS, William R. *Camberley.*
MILLETT, George P. *Scheffenvilli, Canada.*
MISKIN, Leslie T. *Cardiganshire.*
MITCHELL, Thomas W. *Stourport, Worcs.*
MOORE, Roger A. *Salisbury, Wiltshire.*
MUNRO, Flt.-Lt. A. J., R.A.F., *Lincs.*
*NAGESH, Jai Ram. *Ambala, India.*
NENDICK, Michael. *Cheadle, Lancashire.*
NIXON, George. *Newcastle, N. Ireland.*
OLAJIDE, Lawrence O. *Ijebu-Ode, Nigeria.*
OLIVER, Brian M., B.Sc. *Australia.*
ONWUACHU, Clement. *Ilo, Lagos, Nigeria.*
PALMER, Alan. *Northamptonshire.*
PICKARD, Michael J. C. *Wolverhampton.*
RALTON, John Wallace. *Australia.*
REFAULT, Keith. *Channel Isles.*
REINDORP, Peter J. *Letchworth, Herts.*
REYNOLDS, Henry W. W. *Farnborough.*
RICHES, Maurice H. *Chester.*
ROBINS, John K., B.Sc. *London, N.W.2.*
RUSSELL, Rodney. *London, S.E.7.*

SAWTELL, Patrick T. *Bristol.*
SCURRAH, Robert E. *Cleckheaton, Yorks.*
*SOFIZADE, Alex Isaac. *London, N.W.6.*
SPECTOR, Iehuda, B.Sc. *Tel. Aviv.*
STRAY, Ian George. *Surrey.*
TAN KIM HOR. *British North Borneo.*
TAYLOR, Gordon Harbison. *Belfast.*
WALKER, Mervyn Eric. *Aylesbury.*
WEBSTER, Keith. *Norwich.*
WESTMORE, John H. *Isle of Wight.*
WHEELER, A. A. E., D.S.C. *London, S.W.1.*
WIENER, Eliezer. *Magdiel, Israel.*
WIGGINS, John. *London, W.2.*
WINTERBON, James H. *Surrey.*
WOOD, Peter Ian McAskell. *Southampton.*
ZARYWACZ, Shmuel. *Ramal Jan, Israel.*

ADKIN, Francis Whitmore, B.Sc. *Chislehurst.*
ADEBAJO, Adebayo, *Lagos.*
ANANTHARAMAN, Varadarajan, B.Sc. *Kerala State, India.*
BAYLISS, Bryan Howard. *Birmingham.*
BENNETT, Brian L. *Thetford, Norfolk.*
BESWICK, Michael. *Birstall, Leicester.*
BOLAR, Manohar. *Sidcup, Kent.*
CARPENTER, David Harry. *London, W.1.*
CAYZER, John Anthony. *London, S.W.15.*
CHAPMAN, Christopher J. *London, S.E.22.*
CHEN, Soo Peng, B.Sc. *Malaya.*
CHOMPFF, Rudolf Frans E. *N.S.W. Australia.*
CLARKE, Ronald Fambley. *Windsor.*
CLULEY, Brian H. *South Harrow.*
COVERDALE, Francis E. *Oxfordshire.*
COX, David John. *Ruislip.*
CUTLER, William A. *Didcot, Berkshire.*
DADACHANJI, Behrah A., B.Sc. *Bombay.*
DESHPANDE, Shriram, B.Sc. *Bombay.*
DENSHAM, John Kenneth. *Liverpool.*
DHARMAPALA, Chandrasena. *Ceylon.*
DIAS, Ralph Terence, B.Sc. *Bombay.*
DIBBLE, Alan David. *London, S.W.10.*
DINGAR, Surendra Nath, B.Sc. *Lucknow.*
FELDMESSER, Kurt. *St. Albans.*
FENN, Ralph. *Stevenage, Hertfordshire.*
GAETAN, St. Cyr. *Montreal.*
GUPTA, Premshanker Bajranglal, *India.*

* Reinstatements.

U.D.C. 621.382.333

Symmetrical Transistors

By

G. H. PARKS, B.Sc.†

Summary: Significant differences in performance between symmetrical and unsymmetrical alloy junction transistors are revealed by an examination of their properties on a theoretical basis. Brief descriptions are given of some circuits for telecommunication purposes in which the use of symmetrical transistors offers significant improvements over the use of more conventional unsymmetrical transistors.

1. Introduction

A paper by Sziklai[1] published in 1953 described the symmetrical properties of transistors and some of their applications. Two kinds of symmetry were described, one being concerned with the complementary characteristics of *pnp* and *npn* transistors and the other with the symmetry shown by single units in which the emitter and collector may be interchanged. This paper will only be concerned with the latter type of symmetry and the applications of transistors, which have been specially constructed so as to exhibit this symmetry to a high degree, to telephone switching and telephone transmission circuits.

2. Basic Structure

A section through a typical alloy transistor is shown in Fig. 1. Under normal conditions of operation the emitter-base junction is forward biased and the design is such that for a *pnp* transistor the forward current consists almost entirely of holes injected into the base region. These holes then diffuse across the base region to the base-collector junction, which is reverse biased, and are then collected. All the holes injected at the emitter do not, however, reach the collector but some are lost due to recombination with electrons. This takes place to a small extent in the base region but the majority of holes so lost recombine at the surfaces, as shown in the diagram. Conditions are particularly favourable at the surface of the transistor for recombination to take place and considerable effort has been expended in the search for means of reducing this rate of recombination. Apart from this, however, an obvious way of increasing the number of holes collected at the collector is to make the collector area larger than that of the emitter. The section shown in Fig. 1 is thus typical of the very large number of alloy junction transistors being produced at the present time.

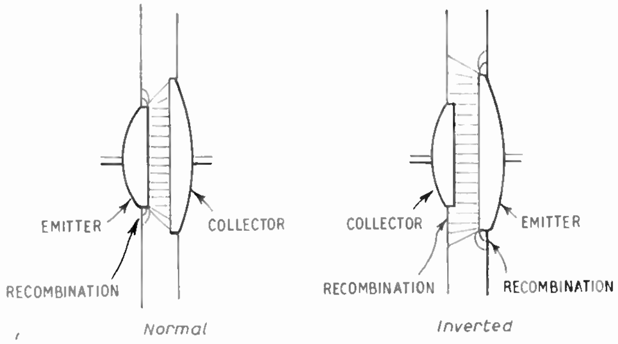


Fig. 1. Unsymmetrical transistor.

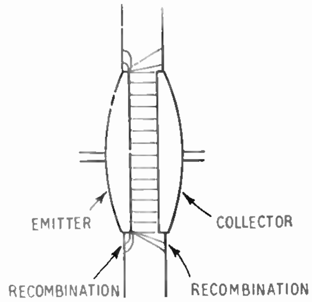


Fig. 2. Symmetrical transistor.

When such a transistor is used in the inverted condition, i.e. with the emitter acting as collector and the collector as emitter, its characteristics are changed appreciably. Thus, as can be seen from Fig. 1, a much greater area is available for recombination to take place, and as would be expected the current gain in this condition is much less. Other characteristics are also changed and a more detailed description of such changes is given in Section 3.

Although the inverted connection is not frequently used the behaviour of a transistor in this condition determines in part its behaviour in the normal but saturated condition. This latter condition, in which both emitter and collector junctions are forward biased but conventional current continues to flow from emitter to collector, is frequently used in switching circuits and a knowledge of the behaviour in this condition is of great importance. The basic relationships are discussed in Sections 3.1.2, 3.1.3 and 3.1.4.

† Associated Electrical Industries (Woolwich) Ltd., Telecommunications Laboratory, Blackheath, London, S.E.

As may be seen from Fig. 2, in a symmetrical transistor, i.e. one in which the emitter and collector areas are equal, a somewhat greater area of surface at which recombination may take place exists compared with that of the unsymmetrical transistor operated in the normal condition. As a result a somewhat lower current gain may be expected. However improvements in surface conditions enable symmetrical transistors to be made which are not greatly inferior in current gain in either direction to the unsymmetrical transistor in the normal condition. The emitter and collector are thus completely interchangeable. In addition other characteristics are modified resulting in superior performance in certain applications.

3. Theoretical Considerations

3.1. *Static Behaviour*

Ebers and Moll[2] have shown that, with certain limitations, analytical expressions can be obtained for *pn* junction transistors from which the gross non-linear behaviour can be determined. These equations are based on the assumed current-voltage relationship for a *pn* junction expressed by the equation

$$I = I_s[\exp(qV/kT) - 1] \qquad \ldots\ldots(1)$$

where k is Boltzmann's constant, q is the electronic charge, T is the absolute temperature, V is the potential across the junction and I_s is the reverse saturation current. As a result the analytical expressions for *pn* junction transistors are given by

$$I_e = I_{es}[\exp(qV_{eb'}/kT) - 1] - \alpha_i I_{cs}[\exp(qV_{cb'}/kT) - 1] \ldots\ldots(2)$$

and

$$I_c = I_{cs}[\exp(qV_{cb'}/kT) - 1] - \alpha_n I_{es}[\exp(qV_{eb'}/kT) - 1] \ldots\ldots(3)$$

where I_{es} and I_{cs} are the reverse saturation currents of the emitter and collector respectively, and α_n and α_i are the normal and inverted short circuit current gain respectively. The equivalent circuit represented by these equations is shown in Fig. 3 where

$$I_{ef} = I_{es}[\exp(qV_{eb'}/kT) - 1]$$

and

$$I_{cf} = I_{cs}[\exp(qV_{cb'}/kT) - 1]$$

The equations (2) and (3) can be further modified to

$$I_e = -\frac{I_{e0}}{1-\alpha_n\alpha_i}\left[\exp\frac{qV_{eb'}}{kT} - 1\right] + \frac{\alpha_i I_{c0}}{1-\alpha_n\alpha_i}\left[\exp\frac{qV_{cb'}}{kT} - 1\right] \ldots\ldots(4)$$

and

$$I_c = -\frac{I_{c0}}{1-\alpha_n\alpha_i}\left[\exp\frac{qV_{cb'}}{kT} - 1\right] + \frac{\alpha_n I_{e0}}{1-\alpha_n\alpha_i}\left[\exp\frac{qV_{eb'}}{kT} - 1\right] \ldots\ldots(5)$$

and the relationship

$$\alpha_i I_{c0} = \alpha_n I_{e0} \qquad \ldots\ldots(6)$$

can be shown to hold, where I_{e0} = saturation current of emitter junction with zero collector current and I_{c0} = saturation current of collector junction with zero emitter current. Let us now consider some of the properties of transistors, and in particular those of symmetrical transistors under different conditions of bias, as illustrated by these theoretical relationships.

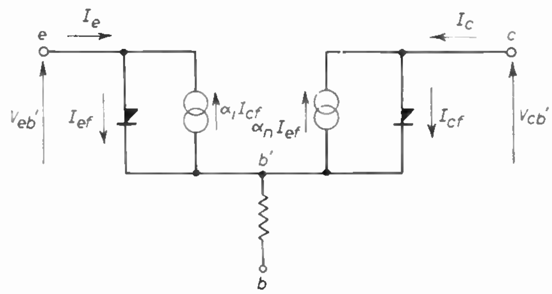


Fig. 3. Equivalent circuit for *pn* junction transistor—static conditions.

3.1.1. Reverse currents flowing when both junctions are reverse biased

It is sufficient that $V_{eb'}$ and $V_{cb'}$ should be large compared with kT/q such that exp $(qV_{eb'}/kT)$ and exp $(qV_{cb'}/kT)$ may be neglected compared with 1. At normal room temperature kT/q = 26 mV. Equations (4) and (5) then become,

$$I_e = +\frac{I_{e0}}{1-\alpha_n\alpha_i} - \frac{\alpha_i I_{c0}}{1-\alpha_n\alpha_i} \qquad \ldots\ldots(7)$$

$$I_c = +\frac{\alpha_n I_{c0}}{1-\alpha_n\alpha_i} - \frac{\alpha_n I_{e0}}{1-\alpha_n\alpha_i} \qquad \ldots\ldots(8)$$

and from these it can be shown that,

$$I_e = \frac{I_{e0}(1-\alpha_n)}{1-\alpha_n\alpha_i} \qquad \ldots\ldots(9)$$

$$I_c = \frac{I_{c0}(1-\alpha_i)}{1-\alpha_n\alpha_i} \qquad \ldots\ldots(10)$$

These currents correspond to those designated $I_{e(sim)}$ and $I_{c(sim)}$ by Chaplin and Owens,[3] i.e. the currents which flow when both emitter and collector are simultaneously reverse biased.

Consider unsymmetrical and symmetrical transistors having the following typical characteristics:

	UNSYMMETRICAL		SYMMETRICAL	
	I_{c0}	$= -5\mu A$	I_{c0}	$= -5\mu A$
	I_{e0}	$= -3{\cdot}6\mu A$	I_{e0}	$= -5\mu A$
	α_n	$= 0{\cdot}98$	α_n	$= 0{\cdot}95$
	α_i	$= 0{\cdot}7$	α_i	$= 0{\cdot}95$
then	$I_{e(sim)}$	$= -0{\cdot}23\mu A$	$I_{e(sim)}$	$= -2{\cdot}5\mu A$
	$I_{c(sim)}$	$= -4{\cdot}7\mu A$	$I_{c(sim)}$	$= -2{\cdot}5\mu A$

Thus we see that the total leakage current is similar, but in the symmetrical transistor it divides equally between the two junctions whereas in the unsymmetrical transistor the collector carries nearly all the current, leaving a very small leakage component flowing to the emitter.

3.1.2. Voltage between collector and emitter when both junctions are forward biased

Two cases must be considered:

(1) common emitter connection;

(2) common collector connection;

(1) Let us assume that the transistor is in the common emitter connection, and that a large negative base current is flowing, sufficient to cause both junctions to be forward biased.

We can obtain from eqns. (4) and (5) the following relationships:

$$\exp(qV_{eb'}/kT) = 1 - \frac{I_e+\alpha_i I_c}{I_{e0}} \qquad \text{......(11)}$$

$$\exp(qV_{cb'}/kT) = 1 - \frac{I_c+\alpha_n I_e}{I_{c0}} \qquad \text{......(12)}$$

and hence

$$V_{ec} = V_{eb'} - V_{cb'}$$
$$= \frac{kT}{q}\log_e \frac{I_e+\alpha_i I_c - I_{e0}}{\alpha_n I_e + I_c - I_{c0}} \cdot \frac{I_{c0}}{I_{e0}} \qquad \text{......(13)}$$

Consider now the case when $I_c = 0$, $I_e \gg I_{e0}$ and $\alpha_n I_e \gg I_{c0}$.

We then have

$$V_{ec(I_c=0)} = \frac{kT}{q}\log_e \frac{I_e}{\alpha_n I_e} \cdot \frac{\alpha_n}{\alpha_i} = \frac{kT}{q}\log_e \frac{1}{\alpha_i} \qquad \text{......(14)}$$

This expression for V_{ec} has been obtained on the assumption that $I_c = 0$, and this may seem a somewhat impractical choice of value for I_c. However, this is the condition which gives the smallest value for V_{ec}, as the collector junction must be sufficiently forward biased to inject holes into the base just to cancel the hole current arriving from the emitter. This is also the point where all curves with $I_e \gg I_{es}$ cross the $I_c = 0$ axis. If a net current is flowing out of the collector then the collector junction has less forward bias and so V_{ec} is greater.

(2) Let us now assume that the transistor is in the common collector connection and that a large negative base current is flowing, sufficient to cause both junctions to be forward biased.

Consider also that $I_e = 0$, $I_c \gg I_{c0}$ and $\alpha_i I_c \gg I_{e0}$ we then have

$$V_{ce} = V_{cb'} - V_{eb'} = \frac{kT}{q}\log_e \frac{1}{\alpha_n} \qquad \text{......(15)}$$

Notice that in this case the current I_c is flowing *into* the collector which means that the transistor is operating in the inverted condition.

For our typical unsymmetrical and symmetrical transistors we have the following results:

UNSYMMETRICAL	SYMMETRICAL
$\alpha_n = 0{\cdot}98$	$\alpha_n = 0{\cdot}95$
$\alpha_i = 0{\cdot}7$	$\alpha_i = 0{\cdot}95$

Normal Operation

$$V_{ec} = \frac{kT}{q}\log_e \frac{1}{0{\cdot}7} = 9\,\text{mV} \qquad V_{ec} = 0{\cdot}75\,\text{mV}$$

Inverted

$$V_{ce} = \frac{kT}{q}\log_e \frac{1}{0{\cdot}98} = 0{\cdot}5\,\text{mV} \qquad V_{ce} = 0{\cdot}75\,\text{mV}$$

The smallest voltage drop between emitter and collector is obtained from the unsymmetrical transistor used in the inverted condition. However, to obtain this low value the gain of the transistor has had to be sacrificed, thus in this case the ratio of emitter current to base current is only 2·3. The symmetrical transistor gives a value for V_{ce} in either direction of operation, which is only slightly greater than that obtained for the inverted operation of an unsymmetrical transistor, whilst still retaining a useful amount of gain. For the values given the ratio of emitter current to base current is 19.

3.1.3. A.c. impedance between emitter and collector

An important property of a transistor used as a switch is the a.c. impedance between emitter and collector with both emitter and collector junctions forward biased. The value of this is obtained by differentiating the expression for V_{ec} with respect to I_c.

Equations (11) and (12) may be rewritten in terms of I_c and I_b using relationship

$$I_e + I_c + I_b = 0 \qquad \text{......(16)}$$

thus $$\exp(qV_{eb'}/kT) = 1 + \frac{I_b + I_c(1-\alpha_i)}{I_{e0}} \qquad \text{......(17)}$$

and $$\exp(qV_{cb'}/kT) = 1 + \frac{\alpha_n I_b - I_c(1-\alpha_n)}{I_{c0}} \qquad \text{......(18)}$$

from which may be obtained

$$V_{ec} = V_{cb'} - V_{cb'} = \frac{kT}{q}\log_e \frac{1+\frac{I_c}{I_b}(1-\alpha_i)}{\alpha_i\left\{1-\frac{I_c}{I_b}\left(\frac{1-\alpha_n}{\alpha_n}\right)\right\}} \qquad \text{......(19)}$$

on the assumption that $[I_b + (1-\alpha_i)I_c] \gg I_{e0}$ and $\{\alpha_n I_b - (1-\alpha_n)I_c\} \gg I_{c0}$

Differentiation of eqn. (19) with respect to I_c, then gives

$$\frac{dV_{ec}}{dI_c}=\frac{kT}{q}\left\{\frac{1-\alpha_n}{\alpha_n\left[I_b-I_c\left(\frac{1-\alpha_n}{\alpha_n}\right)\right]}+\frac{1-\alpha_i}{I_b+I_c(1-\alpha_i)}\right\} \quad (20)$$

Consider now the case when $I_c = 0$:

$$\frac{dV_{ec}}{dI_{c\ (Ic=0)}}=\frac{kT}{qI_b}\left\{\frac{1}{\alpha_n}-\alpha_i\right\} \qquad \ldots\ldots(21)$$

Similarly for the case when $I_e = 0$, it can be shown that

$$\frac{dV_{ec}}{dI_{c\ (Ie=0)}}=\frac{kT}{qI_b}\left\{\frac{1}{\alpha_i}-\alpha_n\right\} \qquad \ldots\ldots(22)$$

Calculation of the a.c. impedance between emitter and collector for the case when I_b = 1mA for a typical symmetrical and a typical unsymmetrical transistor gives the following values:

UNSYMMETRICAL	SYMMETRICAL
$\alpha_n = 0{\cdot}98$	$\alpha_n = 0{\cdot}95$
$\alpha_i = 0{\cdot}7$	$\alpha_i = 0{\cdot}95$
$\frac{dV_{ec}}{dI_{c\ (Ic=0)}} = 8\ \Omega$	$\frac{dV_{ec}}{dI_{c\ (Ic=0)}} = 2{\cdot}5\ \Omega$
$\frac{dV_{ec}}{dI_{c\ (Ie=0)}} = 11{\cdot}25\ \Omega$	$\frac{dV_{ec}}{dI_{c\ (Ie=0)}} = 2{\cdot}5\ \Omega$

Somewhat larger values would be obtained in practice due to the addition of volume resistances but these should be one ohm or less for alloy junction devices. The symmetrical transistor is seen to have an appreciably lower impedance than the unsymmetrical one.

3.1.4. Stored charge in the effective base region

For fast switching purposes the minority carrier charge stored in the effective base region is of considerable importance. If the transistor is allowed to "bottom" in the on condition, that is the two junctions are allowed to become forward biased, this stored charge may be quite large. The result is to slow down the action of switching the transistor both on and off. Thus the transistor cannot complete its switching to the on state until this charge has been delivered from the external circuit, neither can it switch to the off state until this charge has been removed.

Equations (2) and (3) may be written in the following form:

$$I_e = I_{ef}-\alpha_i I_{cf} \qquad \ldots\ldots(23)$$

$$I_c = I_{cf}-\alpha_n I_{ef} \qquad \ldots\ldots(24)$$

where I_{ef} and I_{cf} represents the forward current injected at the emitter and collector respectively.

Consider Fig. 4 which shows the distribution of hole density through a section of the base region with both junctions forward biased.

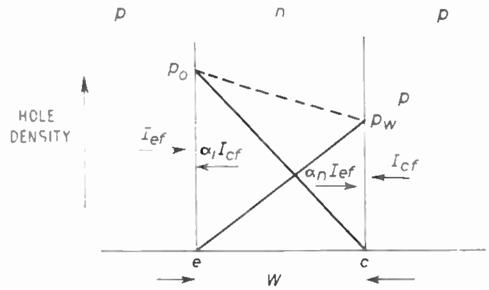


Fig. 4. Distribution of hole density through base region with both junctions forward biased.

The current arriving at the collector from the emitter is

$$\alpha_n I_{ef} = \frac{p_0}{W}\cdot qD_p A \qquad \ldots\ldots(25)$$

where D_p is the diffusion constant for holes and A is the area.

Similarly the current arriving at the emitter from the collector is

$$\alpha_i I_{cf} = \frac{p_W}{W}\cdot qD_p A \qquad \ldots\ldots(26)$$

Hence $$I_e = \frac{qD_pA}{W}\left[\frac{p_0}{\alpha_n}-p_W\right] \qquad \ldots\ldots(27)$$

and $$I_c = \frac{qD_pA}{W}\left[\frac{p_W}{\alpha_i}-p_0\right] \qquad \ldots\ldots(28)$$

Now $$Q_b = qWA.\tfrac{1}{2}(p_0+p_W) \qquad \ldots\ldots(29)$$

where Q_b is the charge stored in the effective base region. From eqns. (27), (28), and (29) the following expression for Q_b is obtained:

$$Q_b = \frac{\tau_b\alpha_n\alpha_i}{1-\alpha_n\alpha_i}\left[I_e\left(1+\frac{1}{\alpha_i}\right)+I_c\left(1+\frac{1}{\alpha_n}\right)\right] \ldots\ldots(30)$$

where $\tau_b = \dfrac{W^2}{2D_p}$

Clearly the base charge increases rapidly as the product of α_n and α_i approaches unity, so that the charge stored in the base of a symmetrical transistor will be greater than that for an unsymmetrical transistor of similar geometry. However, hole storage effects in switching circuits also depend on the time constant associated with Q_b and this we shall see in the following section is smaller for the symmetrical transistor than the unsymmetrical one.

Let us compare once again our two typical transistors, for the case where I_e = 5mA, I_c = −4mA, and hence I_b = − 1mA. This corresponds to a heavily bottomed condition.

UNSYMMETRICAL	SYMMETRICAL
$\alpha_n = 0{\cdot}98$	$\alpha_n = 0{\cdot}95$
$\alpha_i = 0{\cdot}7$	$\alpha_i = 0{\cdot}95$
$Q_b = 8{\cdot}9\tau_b$	$Q_b = 19{\cdot}1\tau_b$

The stored charge for the symmetrical transistor would thus be more than twice as great.

3.2. *Dynamic Behaviour*

3.2.1. Storage time

Moll[4] has shown that in general the storage time for a transistor in the common emitter connection is given by the following expression (with a slight change of notation):

$$t_s = \left(\frac{\tau_n + \tau_i}{1 - \alpha_n \alpha_i}\right) \log_e - \frac{I_{b1} - I_{b2}}{I_c(1-\alpha_n)/\alpha_n - I_{b2}} \quad \ldots\ldots(31)$$

Where τ_n and τ_i are the time constants for the normal and inverted condition and I_{b1} and I_{b2} represent the base current values before and immediately after the beginning of the turn off transient respectively.

$\tau_n = \dfrac{1}{\omega_{\alpha_n}}$ and $\tau_i = \dfrac{1}{\omega_{\alpha_i}}$ where ω_{α_n} and ω_{α_i} are the current gain normal and inverted angular cut-off frequencies respectively.

For an unsymmetrical transistor $\omega_{\alpha_i} \ll \omega_{\alpha_n}$, giving $\tau_i \gg \tau_n$ so that in t_s the factor intrinsic to the transistor is dominated by τ_i and α_i. For a symmetrical transistor, however, $\omega_{\alpha_i} \simeq \omega_{\alpha_n}$ so that $\tau_i \simeq \tau_n$ and on this account the storage time should be reduced. However, the product $\alpha_n\alpha_i$ is now much closer to unity and the expression for t_s is clearly very sensitive to changes in this product. The overall effect is thus likely to be an increase in t_s.

Consider the case in which $I_c = -5$mA, $I_{b1} = -1$mA, $I_{b2} = 0$.

UNSYMMETRICAL	SYMMETRICAL
$\alpha_n = 0{\cdot}98$	$\alpha_n = 0{\cdot}95$
$\alpha_i = 0{\cdot}7$	$\alpha_i = 0{\cdot}95$
$\omega_{\alpha_n} = 2\pi \times 5 \times 10^6$	$\omega_{\alpha_n} = 2\pi \times 2{\cdot}5 \times 10^6$
$\omega_{\alpha_i} = 2\pi \times 1 \times 10^6$	$\omega_{\alpha_i} = 2\pi \times 2{\cdot}5 \times 10^6$
$t_s = 1{\cdot}3$ µsec	$t_s = 1{\cdot}73$ µsec

The expression for the storage time alone involves the current gain and the cut-off frequency in the inverted condition. Both the rise time and fall time involve only parameters relating to the normal condition of operation. Thus in these two cases the only difference in performance between symmetrical and unsymmetrical transistors arises from the fact, that, in general, typical α_n and ω_{α_n} values for symmetrical transistors are somewhat lower than for unsymmetrical transistors of similar geometry.

3.3. *Review of Properties of Symmetrical and Unsymmetrical Transistors on a Theoretical Basis*

We have seen that the symmetrical transistor should possess a number of advantages over the unsymmetrical type as well as some disadvantages. Let us consider what these are and under what circumstances the symmetrical transistor may be expected to have a superior performance.

3.3.1. Leakage current

The leakage currents of a symmetrical transistor may be expected to be similar to those of an unsymmetrical transistor when the opposite electrode is open circuited, but when both junctions are reverse biased the reverse current of the emitter junction of an unsymmetrical transistor may be an order lower and advantage can be taken of this property in, for example, d.c. amplifiers.[3]

3.3.2. Voltage between collector and emitter when both junctions are forward biased

In this condition the voltage between emitter and collector of a symmetrical transistor whether in the normal or inverted mode of operation is only slightly greater than that for an unsymmetrical transistor in the inverted mode and much less than that in normal mode. In addition there is no loss in gain when the symmetrical transistor is operated in the inverted condition. Thus the symmetrical transistor is ideal as a closed switch, being able to handle signals in either direction and of either polarity.[5] Further, although the resistor through which the base current is fed to hold the transistor saturated introduces a shunt loss, this is minimized by the fact that reasonably high gain is obtainable in both directions of operation, enabling a fairly high value of resistor to be used.

3.3.3. A.c. impedance between emitter and collector

With both junctions forward biased, the a.c. impedance between emitter and collector for the symmetrical transistor is likely to be appreciably lower than that for an unsymmetrical transistor.

3.3.4. Stored charge in the effective base region

Equation (30) shows that Q_b, the charge stored in the effective base region is proportional to $1/(1-\alpha_n\alpha_i)$. This factor is clearly very sensitive to the product $\alpha_n\alpha_i$ as it approaches unity so that the stored charge for a symmetrical transistor with values of α_n and α_i very close to unity could become very large. This is a disadvantage in a circuit in which the low voltage drop between emitter and collector, with both junctions forward biased, is being exploited, if rapid switching is also required. This disadvantage is likely to be only partially offset by the fact that $(\tau_n+\tau_i)$ which appears in the expression for the storage time t_s (eqn. (31)) is lower. The reduction in t_s due to this will in general be insufficient to overcome the increase due to the factor $1/(1-\alpha_n\alpha_i)$ which also appears in the expression for t_s.

3.3.5. Summary of properties

The particular merits of a symmetrical transistor are thus:

(1) There is a very low voltage drop between emitter and collector when passing current in either direction with both junctions forward biased.

(2) Reasonable current gain exists between base and either emitter or collector in the saturated condition.

(3) The a.c. impedance between emitter and collector in the saturated condition is very low, and when both junctions are reverse biased is very high.

The symmetrical transistor is thus ideally suited as a bidirectional switch, as it can be made to have a very low insertion loss in the "on" condition and a very high insertion loss in the "off" condition.

4. Applications

4.1. *Channel Modulator*

A typical circuit for a channel modulator is shown in Fig. 5. According to the phase of the carrier supplied to the bases of the two symmetrical transistors, at any instant one is cut-off and the other is on. The signal is thus switched through the low impedance of the "on" transistor, alternately across one half of the winding of the output transformer and then across the other, at a rate determined by the frequency of the carrier supply. The peak voltage of the carrier supplied to the bases of the two transistors must be in excess of the peak signal voltage in order to ensure that the transistor which is off, cannot be turned on by the signal. The base current supplied by the carrier to hold the "on" transistor on, must be sufficient to keep the transistor in the low impedance or "bottomed" condition under the condition of maximum signal current. It should be noticed that symmetrical transistors are necessary in this arrangement as signal current may flow in either direction through each transistor.

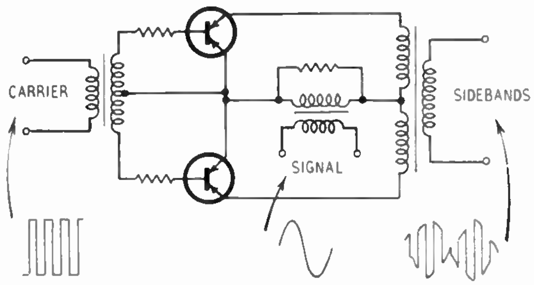


Fig. 5. Circuit and waveforms for a channel modulator.

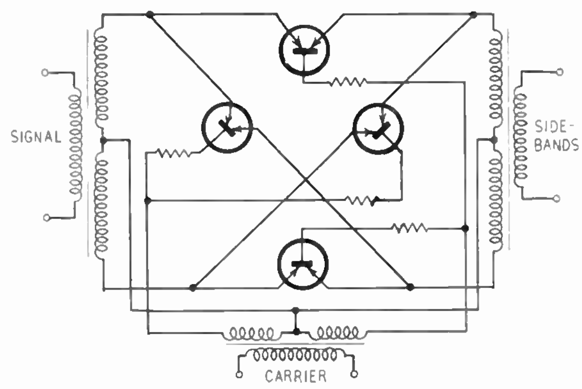


Fig. 6. Circuit of a sub-group modulator.

The advantages of this circuit are that no special selection of transistors is required to obtain a low carrier leak. This is because of the very low voltage-drop across the transistors in the on state and because of the amplification provided by the transistors which enables a low-power carrier supply to be used.

There is a frequency limitation on the circuit, however, dependent on the cut-off frequency of the transistors used. The transistors are bottomed in the on state and this gives rise to hole storage effects. As a result if the attempt is made to switch the transistors at too high a rate, pulses of current at the carrier supply frequency will appear at the output. This is due to the transistor which should have turned off remaining low impedance until all the holes stored in the base have been removed. A calculation of this time, t_s, may be made from eqn. (31) in Section 3.2.1.

4.2. *Sub-group Modulator*

A circuit for a sub-group modulator is shown in Fig. 6. This is similar to the channel modulator except that four transistors are used instead of two. At any instant, two transistors are in the on state and two in the off state. The output waveform is the same as that shown for the circuit of Fig. 5. The signal appears across the whole of the secondary winding of the output transformer first in one phase and then in the opposite phase at a rate determined by the frequency of the carrier supply. A somewhat lower carrier leak is obtainable with this circuit due to the fact that any currents which are passed by the two transistors which are in the process of turning off, tend to cancel in the output.

4.3. *Compressor and Expander Circuits*

In Fig. 9 the a.c. saturation resistance of a symmetrical transistor is plotted against the reciprocal of the d.c. base current. For small signals therefore the collector-emitter path behaves as a low impedance under the control of the d.c. base current. If therefore a circuit such as that shown in Fig. 7 in diagrammatic form is used, a range of output levels less than the range of levels at the input may be obtained.

The base current of the transistor is obtained by rectification of the amplifier output. Thus in the circuit shown, as the base current is increased, the transistor impedance falls. This reduces the input to the amplifier resulting in a compression characteristic; higher input levels are amplified less than lower input levels. With the transistor connected in series with the input, an expansion characteristic can likewise be obtained, i.e. higher input levels are amplified more than lower levels. This characteristic may be made complementary to the compression characteristic.

A signal passing through both compression and expansion circuits which are complementary, has its range of levels unaltered by the process. One advantage obtained by the process, however, is an improvement in signal/noise ratio for noise which is introduced after compression of the signal has taken place. This noise is subject only to the expansion characteristic and being of an assumed low level, is amplified less than the signal, an improved signal/noise ratio thus results.

4.4. *Crosspoint in a Telephone Switching System*

The basic circuit for a bi-lateral switch is shown in Fig. 8(*a*). With the switch in the "off" position, both junctions are reversed biased and the a.c. impedance of each junction is very high, of the order of megohms. Transmission through the switch in either direction is thus subject to considerable attenuation so long as the peak positive input signal does not exceed the supply potential E_1 connected to the base. The equivalent circuit for this condition is shown in Fig. 8(*b*).

Typical values are:

$$R_g = R_L = 600\ \Omega$$
$$R_b = 10\ \text{k}\Omega$$
$$r_e' = r_c' = 4{\cdot}5\ \text{M}\Omega$$

and these would give an insertion loss of more than 120 dB.

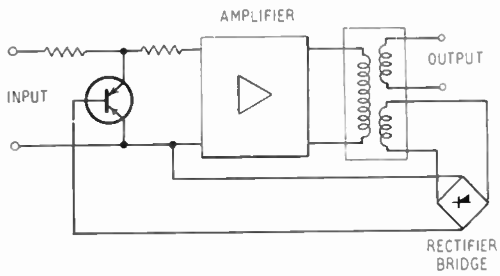


Fig. 7. Compressor circuit.

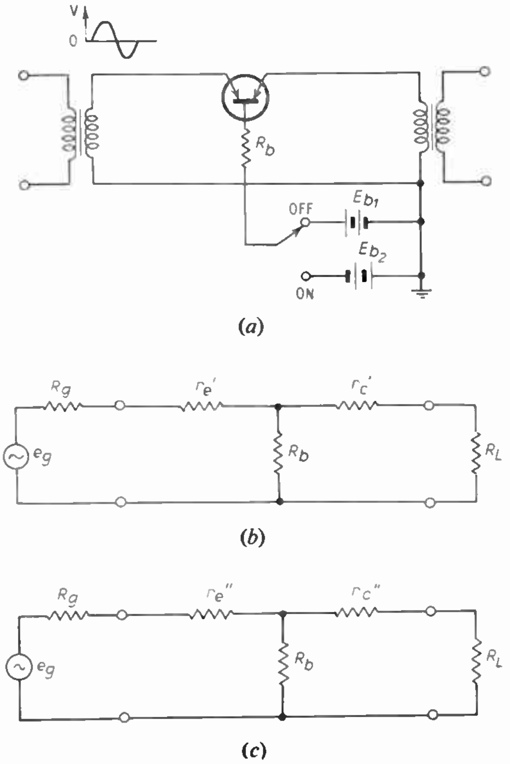


Fig. 8. (*a*) Circuit of the bilateral switch.
(*b*) Equivalent circuit for the "off" position.
(*c*) Equivalent circuit for the "on" position.

With the switch in the "on" position, both junctions are forward biased and both the voltage drop and the a.c. impedance between emitter and collector are low (Sections 3.1.2 and 3.1.3). Transmission is then possible in either direction through the switch with very little attenuation. The equivalent circuit for this condition is shown in Fig. 8(*c*) where typically,

$$R_g = R_L = 600\ \Omega$$
$$R_b = 10\ \text{k}\Omega$$
$$r_e'' = r_c'' = 2\ \Omega$$

resulting in an insertion loss of a small fraction of a decibel. It is interesting to note that in this circuit the signal voltages act as their own d.c. collector-supply potentials.

The circuit is subject to certain limitations governed by the characteristics of the transistor used, these are briefly as follows.

In the "off" condition, the base bias voltage E_{b1} must exceed slightly the peak positive signal voltage e_p. When therefore the signal input is at its peak negative value $-e_p$ the voltage between emitters and base is $e_p + E_{b1} \simeq 2e_p$. As this is a bi-directional circuit it is possible for the input to the collector to be simultaneously at its peak positive value $+e_p$. Thus with the base slightly positive to the most positive of the emitter and collector, twice the peak signal input must not exceed the maximum emitter-to-collector or emitter-to-base voltage of the transistor.

In the "on" condition when the signal voltage is at its positive peak, the base current flowing in R_b is

$$I_b = \frac{e_p + E_{b2}}{R_b} \qquad \ldots\ldots(32)$$

and when at its negative peak, the base current is

$$I_b = \frac{-e_p + E_{b2}}{R_b} \qquad \ldots\ldots(33)$$

If the d.c. current gain of the transistor is the same for both directions of current flow and equal to β, then the maximum signal current that can be allowed to flow with the transistor remaining saturated is

$$I_{\max} = \beta I_b = \beta\left(\frac{-e_p + E_{b2}}{R_b}\right) \qquad \ldots\ldots(34)$$

derived from eqn. (33) as this sets the limit. If the current is allowed to exceed this value when the signal voltage is at its negative peak value, then the transistor will become unsaturated and clipping of the signal will result.

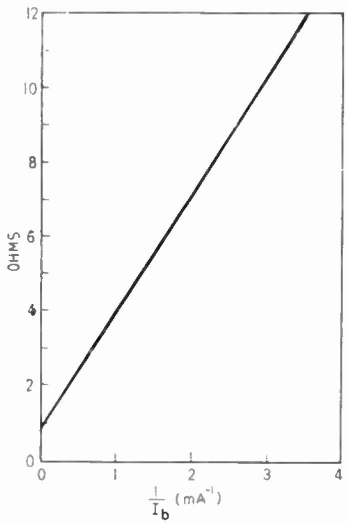


Fig. 9. Saturation resistance of a symmetrical transistor.

The a.c. impedance between emitter and collector of a typical XS 101 transistor in the saturated condition is plotted in Fig. 9.

4.5. *Switch in a Reciprocal Gating System*

Another application is the use of symmetrical transistors in a reciprocal gating system. A low-loss pulse transmission system has been proposed for telephone switching which is also reciprocal.[6, 7, 8] This enables telephone lines to be connected via a storage network and a switch to a common highway, so that communication between any pair of lines can be set up by synchronizing the closing of the switches. These switches can very conveniently be symmetrical transistors controlled by suitable pulses, and the overall loss introduced by the system is theoretically zero.

Figure 10(*a*) shows the elements of such a system. The filter and capacitor in each line form the storage network. It is charged by a substantially smooth flow of power from the signal source and then discharged rapidly into a similar store of the line to which it is connected. This connection is made for the brief period for which the switches in the two lines are simultaneously closed, by the application to their bases of synchronized pulses. The switches are closed for a time $t = \pi\sqrt{(LC)}$ corresponding to the time of one half-cycle of oscillation of the resonant circuit formed by L and C. As a result the current flowing in the switch is a half-sine wave while the voltage across the storage capacitors are half-cosine waves. When the switches are opened at the end of the pulse, the charges of the two capacitors of the two lines in communication will have been completely interchanged with ideally no loss.

To enable this transfer of charge to be achieved in practice with minimum loss, the symmetrical transistors used for the switches must:

(*a*) introduce as little series resistance as possible;

(*b*) introduce as large a shunt resistance as possible;

(*c*) be able to pass the peak current required;

(*d*) be able to withstand peak voltage required.

Let us now consider a practical example. Suppose it is required to handle a peak signal power P_p in an N channel both-way system as shown in Fig. 10(*b*). The peak voltage E_p which appears across the load resistor R_0 is given by

$$\frac{1}{2} \cdot \frac{E_p^2}{R_0} = P_p \qquad \ldots\ldots(35)$$

i.e.

$$E_p = \sqrt{(2P_p R_0)} \qquad \ldots\ldots(36)$$

The peak voltage which appears across the capacitor C_1 is, however, equal to $2E_p$, because the filter, which includes C_1 is unterminated for the period between pulses, i.e. for nearly the complete period. The bias voltage on the base of the transistor must therefore be slightly in excess of this value to hold the transistor cut off. A similar voltage but of opposite sign could simultaneously be present across C_2. Both transistors must therefore be able to withstand a reverse voltage between emitter and base of $2(2E_p) = 4\sqrt{(2P_pR_0)}$.

On the assumption that the period for which the switches are closed $\tau = T/N$ where T is the time between successive closure of the switches then

$$Z_s = R_0 \frac{\tau}{T} = \frac{R_0}{N} \qquad \text{......(37)}$$

In order to determine the peak current which the switch is required to pass, consider the simplified circuit of Fig. 10 (*c*)

$$I_p = \frac{4E_p}{2R_s} = \frac{4E_p}{2R_0/N} \qquad \text{......(38)}$$

$$= \frac{4E\,NP_p}{E_p^2} = \frac{4NP_p}{E_p} \quad \text{from (35)} \quad \text{......(39)}$$

The peak base current required is thus:

$$I_{b(\text{pk})} = \frac{I_p}{\beta} = \frac{4NP_p}{\beta E_p} \qquad \text{......(40)}$$

where β is the common emitter current gain of the transistor at a collector current of I_p and with the transistor just in the saturated condition.

Let P_p = 5mW and N = 25

Then for an XS101 transistor $4E_p$ = 20V.

Therefore from (36)

$$2P_p R_0 = E_p^2 = 25$$

or $$R_0 = \frac{25}{10 \times 10^{-3}} = 2{\cdot}5 \text{ k}\Omega$$

and $$R_s = \frac{R_0}{N} = \frac{2{\cdot}5 \times 10^3}{25} = 100\ \Omega$$

For the transistor

$$Z_s = \sqrt{(Z_{sc} . Z_{oc})} = \sqrt{(R_{\text{sat}} . R_b)} \qquad \text{......(41)}$$

where Z_{sc} and Z_{oc} are the input impedance of the circuit on short circuit and open circuit respectively and R_{sat} is the a.c. resistance of the emitter-collector path.

Hence, from (38)

$$I_p = \frac{4E_p}{2R_s} = \frac{20}{200} = 100 \text{ mA}$$

and $$V_b = 2E_p = 10 \text{ V}$$

Also the peak base current $\dfrac{I_p}{\beta} = \dfrac{100 \times 10^3}{20} = 5\text{mA}.$

For this value of current $R_{\text{sat}} \simeq 2$ ohms (from Fig. 9). Thus, from (41)

$$R_b = \frac{R_s^2}{R_{\text{sat}}} = \frac{10^4}{2} = 5 \text{ k}\Omega$$

The peak base current is also given by

$$\frac{V_p - V_b}{R_b} = I_{b(\text{pk})}$$

We therefore have,

$$V_p - V_b = 25 \text{ V}$$
$$V_p = 35 \text{ V}$$

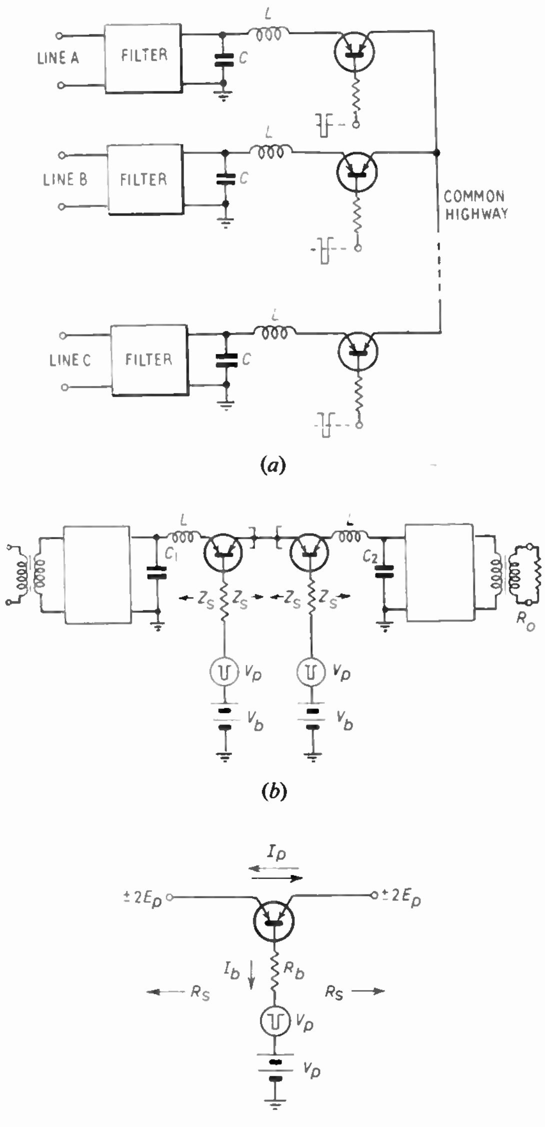


Fig. 10. (*a*) Reciprocal gating system.
(*b*) Circuit of a single channel.
(*c*) Simplified circuit.

Thus using a symmetrical transistor type XS 101 and with R_b = 5 kΩ, V_p = 35 V and V_b = 10 V, the circuit of Fig. 10(*c*) is capable of transmitting a peak audio frequency power of 5 mW, with a theoretical loss of only a fraction of a decibel. In practice due to transformer losses, etc., the loss, in a circuit such as that shown in Fig. 10(*b*) in which two transistors are used per link, will be about 1·5 dB.

5. Conclusions

It has been shown from theoretical considerations how the performance of an alloy junction transistor of symmetrical construction may be expected to differ from that of the more conventional unsymmetrical alloy junction transistor. Symmetrical transistors are seen to offer some advantages in certain circuit applications. This point has been illustrated by brief descriptions of some applications of symmetrical transistors to circuits for telecommunication purposes, in which these advantages are significant.

6. Acknowledgments

Acknowledgment is made to the Director of the Research Laboratory of Associated Electrical Industries (Woolwich) Ltd., for permission to publish this paper. The author also wishes to thank his colleagues in the Transmission Department for information concerning the application of symmetrical transistors to modulator and compandor circuits.

7. References

1. G. C. Sziklai, "Symmetrical properties of transistors and their applications", *Proc. Inst. Radio Engrs*, **41**, pp. 717–24, June 1953.
2. J. J. Ebers and J. L. Moll, "Large-signal behaviour of junction transistors", *Proc. Inst. Radio Engrs*, **42**, pp. 1761–72, December 1954.
3. G. B. B. Chaplin and A. R. Owens, "Some transistor input stages for high-gain d.c. amplifiers", *Proc. Instn Elect. Engrs*, **105B**, pp. 249–57, 1958 (I.E.E. Paper No. 2382M, May 1958).
4. J. L. Moll, "Large-signal transient response of junction transistors", *Proc. Inst. Radio Engrs*, **42**, pp. 1773–84, December 1954.
5. R. B. Trousdale, "The symmetrical transistor as a bilateral switching element", *Trans Amer. Inst. Elect. Engrs*, **75**, pp. 400–3, 1956. *Communications and Electronics*, No. 26, September 1956.
6. H. B. Haard and C. G. Svala, U.S. Patent No. 2718621 : 1953.
7. K. W. Cattermole, "Efficiency and reciprocity in pulse-amplitude modulation. Part I.—Principles", *Proc. Instn Elect. Engrs*, **105B**, pp. 449–62, 1958 (I.E.E. Paper No. 2474R, September 1958).
8. J. A. T. French and D. J. Harding, "An efficient electronic switch—the bothway gate", *Post Office Elect. Engrs, J.*, **52**, Part I, pp. 37–42, April 1959.

Manuscript received 6th May 1960 (Paper No. 605).

U.D.C. 621.396.11.029.62

Reception of B.B.C. Television Sound Transmissions on 41·5 Mc/s at Halley Bay, Antarctica

By

L. W. BARCLAY,
B.Sc. (*Associate Member*)†

Summary: The B.B.C. television sound transmissions from London were monitored at Halley Bay, Antarctica, on 130 days between April and October 1958. The observations show that a large part of the interceptions can be ascribed to normal F layer propagation. In some months propagation was better than expected and this may have been due to the presence of abnormally long hops, though the possibility of some contribution from Es modes cannot be eliminated. Disturbance of the F layer is correlated with absence of signals.

1. Introduction

The exceptional solar activity during the I.G.Y. resulted in the vertical incidence critical frequency of the F2 layer, f_0F2, and the corresponding maximum usable frequency (m.u.f.) at oblique incidence, being abnormally high. Thus ionospheric conditions were favourable for long distance propagation on unusually high frequencies. The observations of f_0F2 at the Royal Society Base at Halley Bay, Antarctica (75°31′S, 26°36′W) described by Bellchambers and Piggott[1] show that very high critical frequencies are obtained at equinoxes even there. In general, the daytime ionization is more intense towards the tropics than at higher latitudes, and hence it is worthwhile to see whether long distance propagation on very high frequencies is practicable over nearly North-South paths. The possibility of reception at Halley Bay of the B.B.C. television sound transmissions from London on 41·5 Mc/s is particularly interesting, both theoretically and from the point of view of possible interference between transmissions in widely separated zones. This frequency was, therefore, monitored in April and May and from July to October 1958, the months in which reception might be possible. Signals were received, using a 5-element Yagi array adjusted for 41·5 Mc/s, on the majority of days in April, May, September and October, during the last two-thirds of August and very occasionally, in short bursts, in July. A test on one day with an oscilloscope clearly showed the frame synchronizing waveform of the vision transmitter despite the aerial arrangements being inefficient at 45 Mc/s.

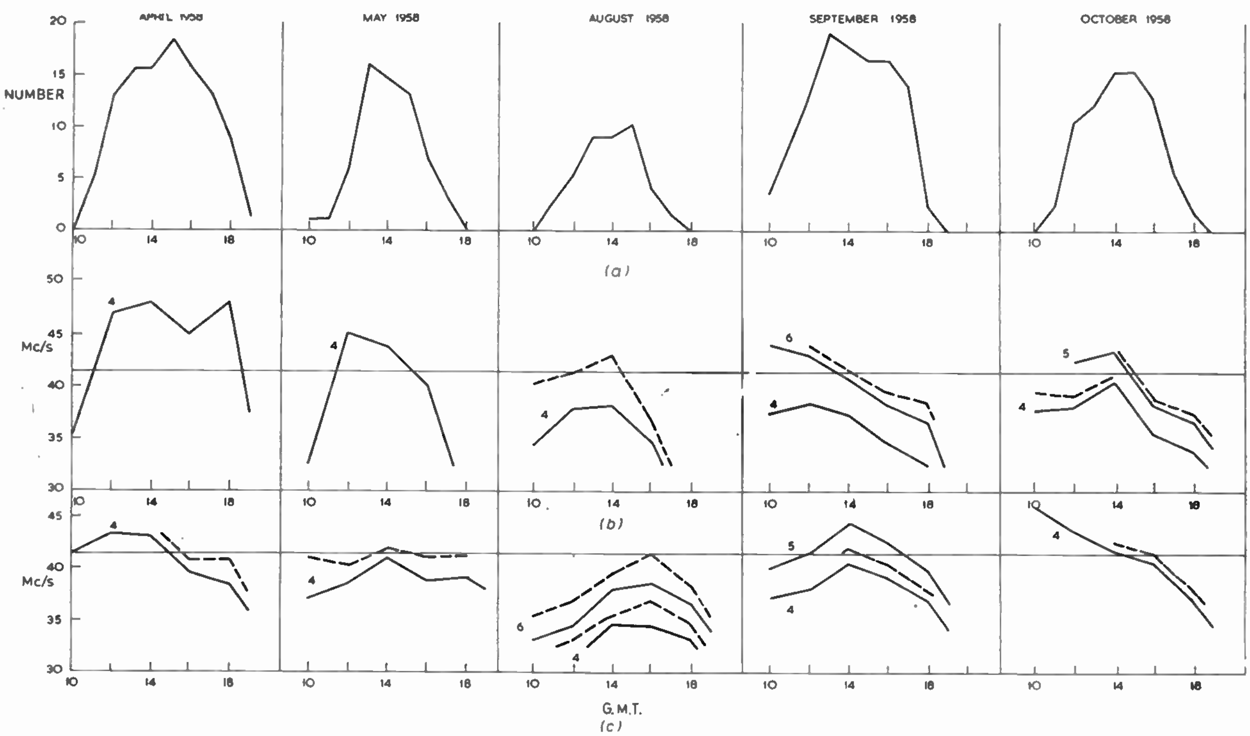
2. Theoretical Considerations

The generally excellent quality and intelligibility of the signals indicated that the propagation was by normal ionospheric reflections rather than by scattering, hence the standard "control point" method may be used for calculating the maximum usable frequencies for the path the length of which is 14 300 km. In this method the m.u.f. for a 4000 km path, MUF (4000), is calculated for control points situated 2000 km along the path from both the transmitter and receiver, and the lower of these computed m.u.f.s is taken to be the m.u.f. of the complete circuit. It is assumed in fact that if the signal can be reflected near the two ends of the path, layer tilts and irregularities will enable the geometry to change so that propagation is possible over the whole path. The critical distance is determined mainly by the lowest angle of elevation at which signals may be transmitted hence, while 4000 km is appropriate on normal h.f. transmission, aerials with great effective heights might give longer first hop distances.

Basically the standard prediction charts prepared by the D.S.I.R. Radio Research Station (Bulletin A, 1958) have been used to compute the m.u.f.s but, as the solar activity differed appreciably from that assumed in preparing the charts, the predicted m.u.f.s have been appropriately modified. The control point m.u.f.s were obtained by taking the predicted MUF(4000) and correcting this by the ratio of the observed to the predicted f_0F2 at a nearby station. Rabat, N. Africa (33°55′N, 6°50′W) was chosen for the Northern control point, while Port Stanley, Falkland Islands (51°42′S, 57°51′W) was used at the Southern end. The corrections were applied at corresponding local times, Port Stanley being 45° farther West than the control point. In addition the MUF(5000) and MUF(6000) values were also computed for certain months by taking the predicted MUF(4000) at 2500 km or 3000 km from the ends of the path, correcting as above, and then allowing for the higher m.u.f. factor (the ratio of m.u.f. to f_0F2) at the increased ranges.

If the median m.u.f. at the control point with the lower absolute value were equal to the working frequency, 41·5 Mc/s, it would be expected that

† The Royal Society I.G.Y. Antarctic Expedition, c/o Burlington House, Piccadilly, London, W.1; now with Marconi's W. T. Co. Ltd., Baddow Research Laboratories, Chelmsford.



——— median values – – – – upper quartile values. *Note:* M.u.f. curves marked 4 refer to 4000 km, 5 to 5000 km and 6 to 6000 km.

Fig. 1. Comparison of the incidence of 41·5 Mc/s signals from London at Halley Bay with the m.u.f.s at the Southern and Northern control points.

(*a*) Occurrences of signal reception in each month. (*b*) M.u.f.s at Southern control point. (*c*) M.u.f.s at Northern control point.

signals would be received on half the days of the month. The upper quartile of the m.u.f. distribution, estimated from the distribution of f_0F2, should correspond to about 7 days interception per month.

The effective rate of interception at each hour may, therefore, be interpreted in terms of the m.u.f.s for the practical circuit and thus the observed and computed m.u.f.s may be compared.

3. Comparison of Number of Interceptions of 41·5 Mc/s at Halley Bay and the Predicted M.U.F.

In Fig. 1 the number of interceptions of the 41·5 Mc/s signal is plotted against the time of day in G.M.T. for each month, together with the median m.u.f. values for the control points. The m.u.f. curves are labelled by a figure denoting the first hop distance in thousands of kilometres.

On some days it was not possible to monitor the signal frequency, and these have been allowed for by multiplying the actual number of interceptions by the ratio of the total number of days in the month to the number of days on which observations were made. These corrections are usually small, as can be seen by the distribution of "C"s (days of no observations) in the day-to-day plot given in Fig. 2. In all 33 "C"s occur in the 5 months of observations.

It must be noted that the B.B.C. test transmission did not usually begin before 1000 G.M.T. and that only normal programmes, starting near 1200 G.M.T. were radiated on Sundays. Before 20th April and after 4th October the test transmissions usually started an hour later. Thus the shortage of pre-noon interceptions in April, May and October, when the m.u.f.s were high in the morning period, was probably almost entirely due to the absence of transmissions.

Turning to the detailed comparisons for the periods of most frequent interception, it is clear that the predicted values of MUF(4000) were low by roughly the amounts shown in Table 1.

Table 1

Increase in theoretical MUF(4000)F2 necessary to account for the observed propagation

MONTH	NORTH END	SOUTH END
April	3%	—
May	3%	—
August	15%	3%
(Upper quartile values)		
September	5%	15%
October	—	7%

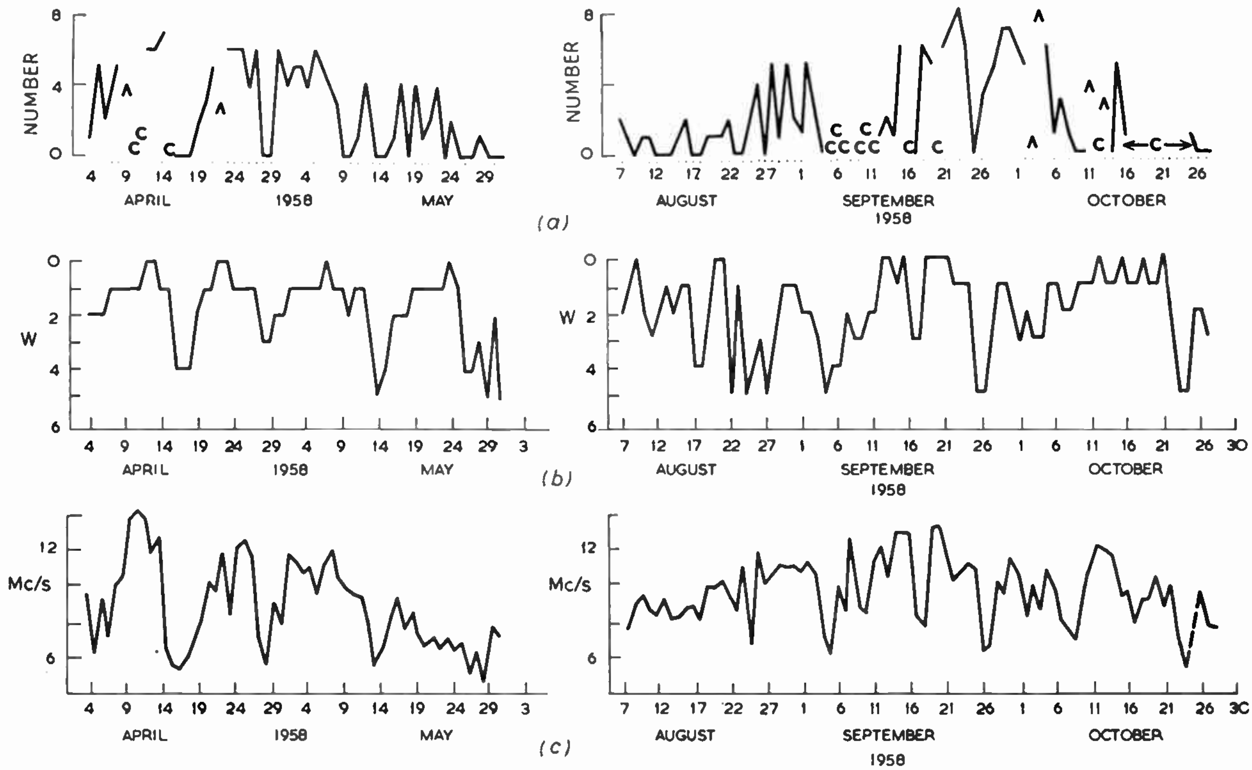


Fig. 2. Comparison between day-to-day changes in the incidence of reception of 41·5 Mc/s with certain F2 layer parameters—April to May 1958.

(*a*) Number of hours in which a signal was detected on each day. "C" means equipment not in use.
(*b*) World-wide F2 layer disturbance index. (*c*) Noon-values of f_0F2 at Halley Bay.

However, it should be remembered that the computed m.u.f. factors would be expected to be slightly low both because of the systematic inaccuracies in reducing ionograms[2] and because the earth's magnetic field tends, on the average, to make them appear lower than the true values. These phenomena may amount to 5–7% in practice. Thus the agreement is excellent in April, May and October but the circumstances present in August and September need to be reconsidered.

It is instructive to consider whether the discrepancies could be explained in terms of abnormally long hops. Evidence that very long hops can occur has been given by Wilkins and Shearman[3] and by Warren and Hagg.[4] Such long hops imply, in the absence of systematic ionospheric tilt phenomena, that the angles of elevation of the signals at the transmitter, receiver, or both, as appropriate, are very low. For large signals to be received, it is necessary for the aerials to be efficient at these angles, i.e. for the effective height above ground to be large. At the Crystal Palace transmitter the aerial is centred 430 ft (approximately 18 wavelengths) above ground level,[5] while at Halley Bay the receiving aerial was centred 10 ft above the dry snow surface of the ice shelf, which extended for several miles in the direction of the transmitter, with the sea approximately 500 ft below. Thus the high effective aerial heights make such long paths possible.

Increasing the hop length to 5000 km or 6000 km increases the m.u.f. factor by 4 to 7% and the increase in critical frequency towards the equator is also between 3 and 10%, varying with month. The computed curves are shown in Fig. 1. Clearly the August, September and October data require very long hops at one or both ends in order to explain the observed circuit behaviour by F propagation alone. Allowing for the systematic errors mentioned before, however, the observations agree reasonably with the predictions using this mode alone. It should, however, be noted that sporadic-E layer phenomena may be contributing to the propagation.

4. Comparison of Day-to-day Changes in Oblique Propagation and F Layer Characteristics

It is interesting to investigate whether the day-to-day changes in the propagation of the 41·5 Mc/s waves are connected with variations in the m.u.f. of the F2 layer and attempts have therefore been made to relate this propagation to indices based on F layer characteristics. As F2 layer disturbance increases near the auroral zone, the observation of f_0F2 at

Halley Bay should provide a very good index of disturbance in the Southern section of the path.

For the path as a whole the world-wide F2 layer disturbance index W prepared by Piggott,[6] which is based on a wide range of F2 layer disturbance phenomena, has been used. This index applies to the whole 24 hours and takes account of the many F2 layer disturbances which are confined to limited parts of the world. Hence an exact relationship would not be expected for the London-Halley Bay path which is open for only a few hours of the day. However, the largest disturbance values represent widespread F2 layer perturbations and might be expected to be correlated with this propagation.

Using the 120 days on which systematic observations were made, tests have been made to determine whether these two indices show correlations with the observations. Figure 2 shows the number of hours of reception for each day together with the W-index and the mean f_0F2 at Halley Bay for the three hours around local noon. On days marked "C" the 41·5 Mc/s channel was not monitored. Examining these data, it is found that the correlation coefficient between signal reception and the W-index is $-0{\cdot}40$, while that between reception and Halley Bay noon values of f_0F2 is $+0{\cdot}56$. The probability of obtaining these values for the correlation coefficient by chance is less than 0·1%. Thus disturbed conditions in the F layer are likely to disrupt this propagation and high values of m.u.f. near Halley Bay are likely to be associated with good propagation. Neither of these phenomena is likely to be correlated with equatorial sporadic E.

It is instructive to examine the number of days with abnormally good or bad propagation when the two indices are unfavourable. The results are shown in Table 2. These figures have been subjected to χ^2-tests for significance and the resultant probabilities of chance occurrence are given in the table.

Similar values, very significantly different from those expected by random sampling, may be obtained by examining days with high values of f_0F2 at Halley Bay when the period of reception was unusually long. For example, when $f_0\text{F2} \geqslant 10$ Mc/s, the signal was heard for 6 or more hours on 16 days out of a possible 19. This also leads to a probability of random occurrence of less than 0·01%. The W-index, however, does not give significant results on days of very good reception. It is apparent from Fig. 2 that both signal reception and Halley Bay f_0F2 values show a seasonal term, peaking just after each equinox, whereas, of course, the disturbance index does not. This is probably the reason for the results just discussed.

Thus we may conclude that both high values of m.u.f. in the South and ionospherically quiet conditions are necessary for propagation and that the amount of ionization at the Southern end of the path is a factor governing the exact period of 41·5 Mc/s propagation.

5. Acknowledgments

The assistance of Messrs. P. M. Brenan, W. H. Bellchambers and D. L. M. Cansfield in maintaining the routine observations at Halley Bay is acknowledged. Helpful advice in the preparation of this paper was given by Mr. W. R. Piggott. This analysis was carried out at the Radio Research Station, D.S.I.R., Slough, by kind permission of the Director.

6. References

1. W. H. Bellchambers and W. R. Piggott, "Ionospheric measurements made at Halley Bay", *Nature (London)*, **182**, pp. 1596–7, 6th December 1958.
2. A. J. Lyon, and A. J. G. Moorat, "Accurate height measurements using an ionospheric recorder", *J. Atmos. Terrest. Phys.*, **8**, pp. 309–17, June 1956.
3. A. F. Wilkins, and E. D. R. Shearman, "Back-scatter sounding: an aid to radio propagation studies", *J.Brit.I.R.E.*, **17**, pp. 601–16, November 1957.
4. E. Warren and E. L. Hagg, "Single-hop propagation of radio waves to a distance of 5,300 km", *Nature*, **181**, pp. 34–5, 4th January 1958.
5. F. C. McLean, A. N. Thomas and R. A. Rowden, "The Crystal Palace television transmitting station", *Proc. Instn Elect. Engrs*, **103B**, pp. 633–43, September 1956.
6. W. R. Piggott, *Proc. U.R.S.I.-A.G.I. Meeting, Brussels*, 1959.

Manuscript received 9th June 1960 (Contribution No. 29)

Table 2

Propagation of 41·5 Mc/s to Halley Bay when (*a*) f_0F2 at Halley Bay $\leqslant$ 8Mc/s and (*b*) F2 layer disturbance index $W \geqslant 3$

Class of days	Total no. of days in class	(*a*) For days with $f_0\text{F2} \leqslant 8$ Mc/s			(*b*) For days with W index $\geqslant 3$		
		actual number found	number expected by chance	probability of this actual occurrence by chance	actual number found	number expected by chance	probability of this actual occurence by chance
No reception	34	24	13	<0·01%	19	9	<0·01%
$\geqslant$ 4 hours reception	48	5	18	<0·01%	5	13	<0·1%

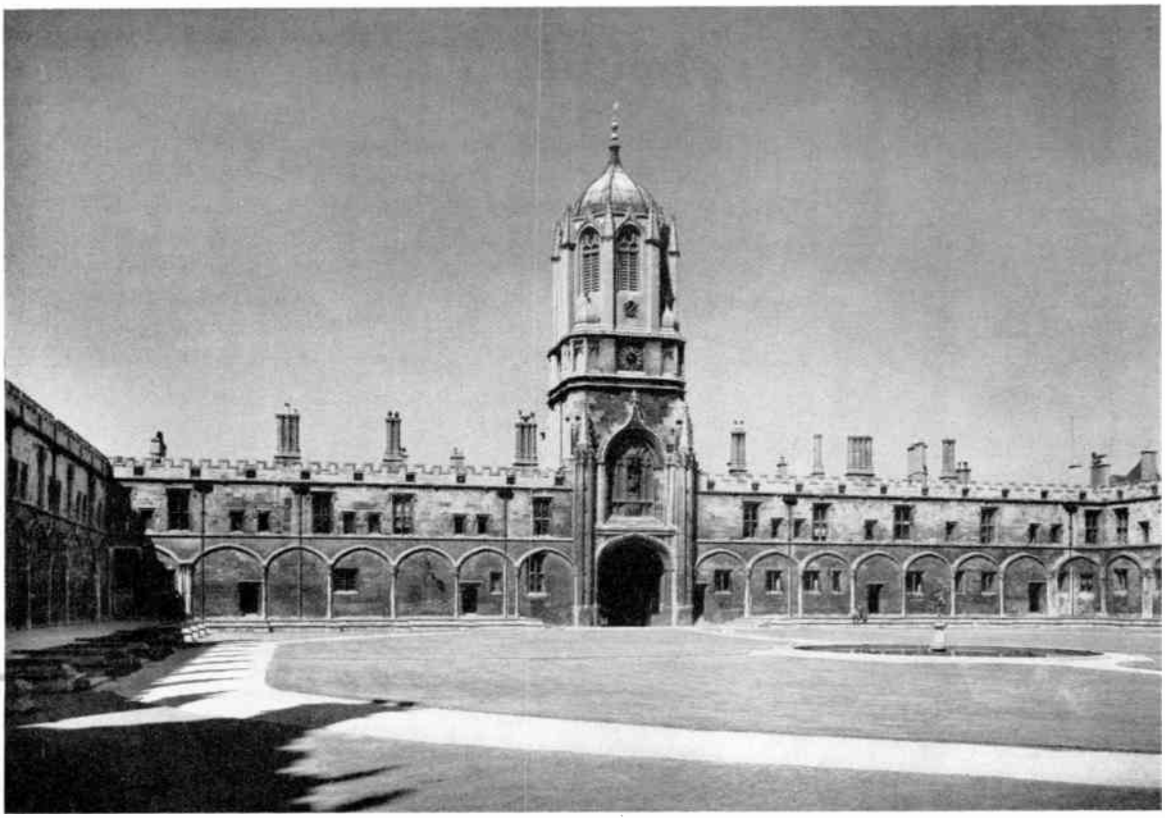
"Radio Techniques and Space Research"

THE 1961 Brit.I.R.E. CONVENTION

The announcement last July that the 1961 Convention will deal with "Radio Techniques and Space Research" has aroused great international interest. Since the launching of the first earth satellites in 1957 several countries have encouraged increasing engineering effort in the design and construction of equipment for measuring physical conditions outside the earth's atmosphere and indeed far into the solar system. An aspect of immediate practical application was emphasized with the launching last year by the United States of America of both passive and active relay satellites for the purpose of linking far apart points on the surface of the earth via outer space.

Until now the United States of America and Soviet Russia have been the principal originators of space research activities, but an increasing programme of work is being undertaken in Great Britain, the Commonwealth, and other European countries with the ultimate object of stepping up the number of research investigations into outer space and the placing of communications satellites in orbit. This latter project in particular has been widely discussed in the House of Commons† and Mr. Peter Thorneycroft, the Minister of Aviation, has also taken part in discussions with the Commonwealth and some European countries on a combined effort. The idea of such a "Commonwealth Club" was suggested by the Institution as long ago as last July in an editorial article on "Space Research". A conference of European Governments is taking place at the end of January in Strasbourg to study the technical and financial possibility of developing large rocket boosters for launching satellites. It is expected that in addition to Great Britain and France, Government representatives from Belgium, Denmark, Federal Germany, Holland, Italy, Norway, Spain, Sweden, Switzerland and Austria will also attend the conference.

† See *Hansard* for 21st December 1960.



Christ Church, Oxford.

Political, economic and financial considerations are involved in a project which will call for a capital expenditure of certainly not less than some hundreds of millions of pounds. In order to show the scientific and engineering development which a space programme demands the Institution has decided to provide a forum for the discussion of the engineering aspects of space research and communication. Radio and electronic engineers from all over the world have therefore been invited to Oxford in July to present and discuss papers on all aspects of this subject which is so challenging to their ingenuity.

As has already been stated in the *Journal*, the 1961 Convention Committee has been at work for several months. The Chairman is Mr. Ieuan Maddock, O.B.E., B.Sc., who is a member of the Brit.I.R.E. Council and represents the Atomic Energy Authority on the British National Committee on Space Research. The Committee's first task has been to draw up a provisional programme of sessions, and papers are being invited for inclusion under the following headings:

Review of the Role of Radio Techniques in Space Research.
Extra-Terrestrial Measurements.
Telemetry.
Techniques in Radio Astronomy
Satellite Engineering Problems.
Communication Satellites.

These headings indicate that it is intended to cover fully all aspects of space research and communications which involve the radio and electronic engineer. It is expected that over the period of four days of the Convention—July 5th to 9th—about 50 papers will be presented.

Submission of Papers

Members who are working in the field covered by the theme of the Convention are invited by the Committee to submit papers for consideration for inclusion in the programme. Initially a synopsis of up to 200 or 300 words should be sent to the Institution and it would be helpful if an idea of the estimated length of the proposed paper could be given.

The Convention Committee also invite readers of the *Journal*, as well as members, to submit either synopses of proposed papers, or to advise the Committee of authors/organizations who might sponsor papers.

The Fifth Clerk Maxwell Memorial Lecture

It is now a tradition of the Institution for the Clerk Maxwell Memorial Lecture to be given by an eminent scientist at a Convention, and this practice will be followed in Oxford this year. The subject of the Lecture will be very relevant to the theme of the Convention and details will be announced in due course.

The 1961 Convention Venue

Members will recall that Oxford was the venue of the 1954 Convention on "Industrial Electronics" and that the headquarters were in Christ Church. The rebuilding programme which was in hand in 1954 has now been completed, and it will be possible to accommodate 300 delegates in Christ Church. Additional accommodation will be available in other colleges. These facilities will be restricted according to the capacity of the lecture theatres.

Official Banquet

It is also hoped that the increased capacity of Christ Church Hall will enable all delegates to attend the Banquet which is being held during the Convention. This will be the second major social function associated with the Convention because four weeks previously, on 8th June, an Institution Dinner will be held in London at the Savoy Hotel and this will have as its main purpose the "launching" of the Convention.

As far as the radio and electronic engineer is concerned, the Clarendon Laboratory is the hub of the scientific work carried out at the University of Oxford. All the technical meetings of the Convention will be held in the large lecture theatre of the Laboratory which is now known as the Lindemann Lecture Room. Use will also be made of the Townsend Lecture Room for overflow purposes, and arrangements will be made for the two theatres to be linked by closed-circuit vision and sound. In many cases the papers will be supported by demonstrations of equipment and these will generally take place elsewhere in the Laboratory. One activity for which the Clarendon Laboratory is world-famous is low temperature physics and the Administrator of the Laboratory, Dr. A. J. Croft, is kindly arranging for delegates to see some of the work which is being carried out in this field.

Further information about the Convention arrangements will be published each month in the *Journal*, including titles and synopses of papers accepted for publication. Registration forms will shortly be circulated, but all members—especially those overseas—who intend to attend the Convention are invited to notify the Institution and make provisional reservations. The cost of attendance, preprints, and residence in Christ Church for the whole period of the Convention will be £18 10s. to *members only*.

The admission of non-members will be dependent on the provisional registrations of members and for this reason it is earnestly requested that members should place reservations NOW—if only on a provisional basis.

U.D.C. 621.373.42:621.396

The Problem of Frequency Synthesis

By

H. J. FINDEN †

Introductory paper at a Symposium on Stable Frequency Generation held in London on 25th May, 1960.

Summary: Frequency synthesis employs precise frequency standards to obtain by means of a frequency conversion process many output frequencies from the one accurate source, each of which can be individually selected. The output frequencies are multiples of harmonic fractions of the frequency standard.

The paper reviews the basic techniques of addition or subtraction and division employed and analyses some of the present frequency synthesizer designs.

1. Introduction

In recent years there has been an enormous increase in the use of radio for communication and navigation. This rapid expansion has put a premium on frequency accuracy in order to use every available channel in a given frequency spectrum. Military communication requirements necessitate rapid access to any channel and the logical channel assignment is on a modular basis[1] e.g. 1 kc/s or 4 kc/s steps. Hence in the h.f. spectrum, 2–27 Mc/s, it may be necessary to generate 25 000 channels and it is because of this that separate precision crystals for each channel become impractical. For a single-sideband suppressed carrier method of operation these frequencies need to be available at the receiver to a very high accuracy of frequency precision in relation to the carrier frequency.

Until recently the most precise standards which were available have been developed at a few frequencies, such as at 100 kc/s and 5 Mc/s, using quartz controlled oscillators. Atomic and nuclear resonance reference standards have now become a practical alternative to the quartz oscillator and developments are in hand to obtain even higher precisions. To take advantage of these precise frequency standards, frequency conversion techniques have been established whereby any frequency can be obtained for a communication process. These techniques, which are the concern of this Symposium, are considered under the term frequency synthesis. Originally the frequency synthesizer was developed for measurement purposes, but the needs of modern communication and navigation systems have provided fresh incentives to devise practical frequency conversion techniques.

Frequency synthesis may be defined as a frequency conversion process to obtain any one of many input frequencies with the same frequency precision as a single frequency standard. A strict definition would state that any of the many possible frequencies must be a multiple of a harmonic fraction of a single frequency standard where the harmonic number of the output frequency can be precisely defined.

A device that produces any one of the frequencies obtained from this frequency conversion process has been called a "frequency synthesizer", e.g. from a 100 kc/s frequency standard a 1 kc/s source can be derived and every multiple of this frequency up to 100 Mc/s generated, i.e. 100 000 individual frequencies.[2]

Many of the existing synthesizer designs use upwards of eighty valves or transistors with mixing techniques for the frequency conversion process. This true synthesis process requires complex filters but has the advantage over other methods of frequency generation in that the resultant output is precisely the same as the frequency standard. Unfortunately, even with a careful choice of the frequencies applied to the mixers, low level spurious frequencies nearly coincident to the wanted signal are generated. Even with complex filters it is not possible to suppress these spurious frequencies to a level more than about 70 dB below the desired signal output.

An alternative approach to the pure synthesis technique is to derive the required frequency from a locked oscillator by a partial synthesis process. The locked oscillator technique makes it possible to achieve a substantially pure output and in practice the problem is that of screening the various multiplier and divider sources from the final desired output. A reference technique which allows the oscillator to be phase-locked by an alternative counting technique will permit sufficient freedom from jitter to be achieved for many purposes with a considerable reduction in the complexity of the equipment.

The necessity for a pure output is particularly

† The Plessey Co., Ltd., Electronic Research Laboratories, Roke Manor, Romsey, Hampshire.

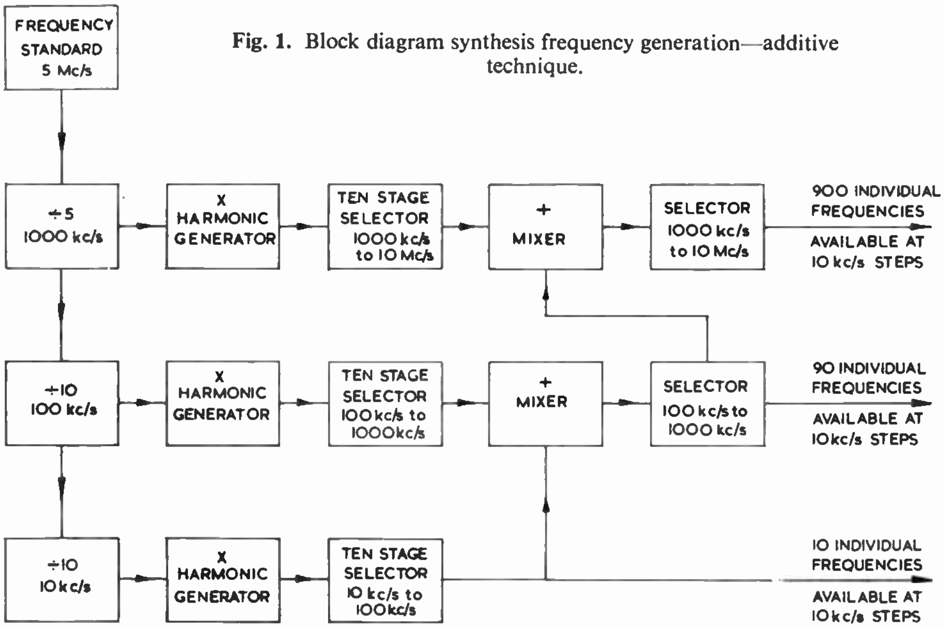


Fig. 1. Block diagram synthesis frequency generation—additive technique.

stringent for the receiver in a communication system where the spurious signal level should be at least 100 dB below the wanted signal. It is now practical to apply digital techniques to military systems where reliability, light weight and compact equipments are essential. These requirements are best achieved by designing to satisfy a particular requirement.

Three basic techniques can be used in the design of precision frequency generators controlled by single frequency standards.

1.1. *Classical Synthesis or Addition Method*

Figure 1 shows a simplified block diagram for a pure synthesis technique.

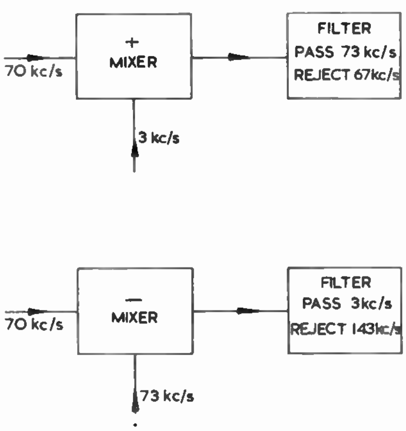


Fig. 3. Simplified filter requirements for synthesis and analysis techniques.

A frequency standard of 5 Mc/s is reduced by a division circuit to provide a signal at 1 Mc/s. Two further division circuits provide signals at 100 kc/s and 10 kc/s. Three multiplier circuits generate harmonics of the decaded input frequencies and are followed by circuits for selecting harmonics up to the tenth order. Hence thirty frequencies controlled in accuracy by the frequency standard are made available for the synthesis process. Two addition circuits followed by selectors permit one thousand frequencies to be synthesized at 10 kc/s steps in the frequency spectrum 10 kc/s—10 Mc/s. As will be described later, the main problems are in the design of efficient selectors and summing circuits. In particular low level spurious frequencies are generated in the summing circuit and the cross-modulation frequencies produced may become nearly coincident with a wanted product and hence present an impossible selection or filter problem.

1.2. *Analysis or Subtractive Method*

The antithesis of the synthesis technique is essentially a subtractive one. Figure 2 shows a simplified block diagram of this method and it will be seen that many of the elements are similar to the synthesis block schematic. Thus the frequency standard has a division chain which feeds chains of multipliers and selectors. However, the additive circuits are replaced by subtractive elements. Each decade stage has the function of subtracting two definite frequencies commencing with the subtraction of the highest frequency decade harmonic from the frequency of the slave oscillator. The final difference frequency is

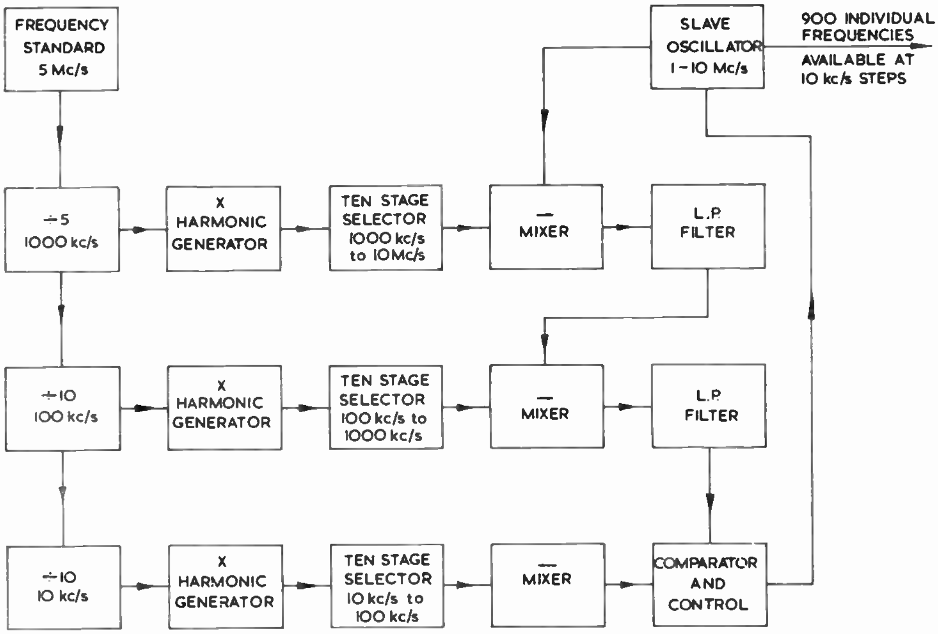


Fig. 2. Block diagram analysis frequency generation—subtraction technique.

compared with the lowest decade harmonic in a comparator which is used to control the frequency of the slave oscillator.

From Fig. 3 it will be seen that in general the filter requirements are easier for the analysis method.

1.3. *Division Method*

Figure 4 shows a block schematic for a division technique. A frequency standard feeds a division chain and as shown is divided by 5, 10, 10 and 10 to provide a 1 kc/s reference. The frequency of a slave oscillator shown as covering the spectrum from 1000 kc/s to 100 kc/s can be divided by any integer between 1000 and 100 and the frequency so obtained is compared with the 1 kc/s reference. A servo loop is used to control the frequency of the slave oscillator which can be locked in any 1-kc/s step in the range 1000 kc/s—100 kc/s, i.e. any of 901 combinations.

This method depends on digital techniques and in general results in more compact circuits than the synthesis and analysis method.

In order to examine the relative complexity of the techniques employed for precision frequency generation, the various methods available for carrying out the functions required will now be discussed.

2. Frequency Dividers

Frequency dividers may be considered under two general groups. The first group give a signal output even though there is no signal input. The second group are those types of divider which do not give an output signal unless there is a signal input. Obviously the second group are preferable although in general the first group are more economical to manufacture.

Example of first group:

$\left.\begin{array}{l}\text{Multivibrators}\\\text{Phantastrons}\\\text{Blocking oscillators}\end{array}\right\}$ Relaxation circuits

All these circuits have a natural period dependent on their time constant. Division is obtained by

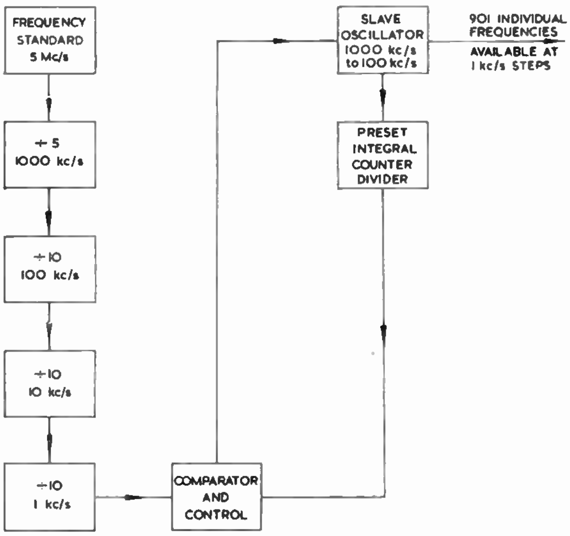


Fig. 4. Block diagram counter divider frequency generation—division technique.

injecting the signal to be divided in such a way as to alter the natural period of the device to be in synchronism with every nth pulse where the division ratio required is n. All these circuits are affected by variations of the operating voltages and valve characteristics and incorrect division ratios can occur fortuitously. Tuned circuits can be employed to indicate an incorrect division ratio.[3] These relaxation circuits are well known and will not be discussed further.

Examples of second group:

Parametric sub-harmonic oscillators.
Regenerative modulator frequency divider.
Counters.

The first two of these, parametric oscillators and regenerative modulators, require tuned circuits, and thus the divider is limited to a particular frequency ratio. Counters, however, basically have aperiodic circuits and any one circuit can be used from some maximum frequency range down to d.c.

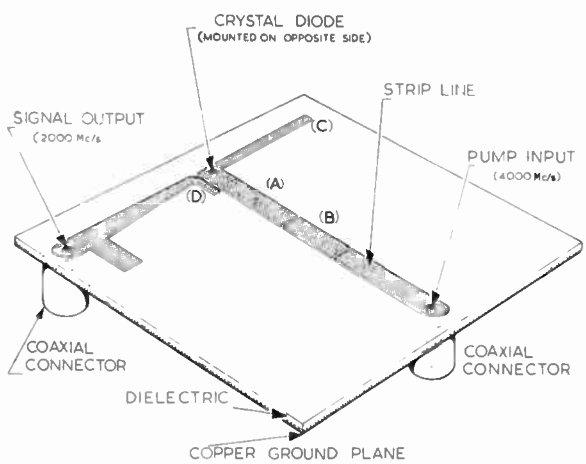


Fig. 5. Stripline subharmonic oscillator.

All those three general types of divider are in use today and the type chosen depends on the particular application. Examples with brief assessments of their properties follow.

2.1. *Parametric Sub-harmonic Oscillator*

The advantages of this type of divider are in the relative simplicity of their circuits and their ability to operate in the u.h.f. range. Stertzer and Bacon[4] described a sub-harmonic oscillator which operated at 2000 Mc/s with an input frequency of 4000 Mc/s. The frequency of operation is limited primarily by the diode used as the non-linear capacitance but it is probable that the maximum operating frequency can be raised by an order of magnitude.

The sub-harmonic oscillator was constructed in stripline form as shown in Fig. 5. The important components are the 2000 Mc/s half-wave resonator with the semi-conductor diode mounted at one end (A); a 4000 Mc/s half-wave resonant strip which permits the 4000 Mc/s drive, sometimes called the pump frequency, to enter resonator but prevents 2000 Mc/s oscillator power from returning (B); a d.c. return which allows the correct bias to be placed in the diode (C); and a loosely coupled output arm containing a pump-regeneration stub (D). The 4000 Mc/s pump causes a negative resistance to be established across the diode and when the losses in the 2000 Mc/s circuit are low enough the circuit will oscillate at that frequency. Negative resistances can also be developed at other sub-harmonics of the pump frequency.

2.2. *Regenerative Dividers*

The limitations of the parametric sub-harmonic oscillator's very narrow bandwidth (approximately 1 %) and the difficulty of using sub-harmonics other than the second can be partially overcome by using regenerative dividers as described by Miller.[5] A generalized block diagram of such a divider is shown in Fig. 6. The circuit can become oscillatory when the

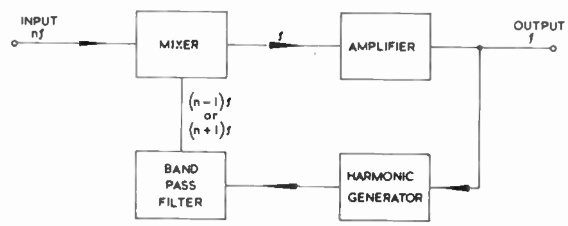


Fig. 6. Generalized regenerative divider ratio n:1.

frequencies are as shown. If an output frequency f is required from an input frequency nf a harmonic generator and filter combination derives the $(n-1)$ or $(n+1)$ harmonic of the output frequency. This harmonic is mixed with the input frequency to produce f and the feedback loop is completed.

This technique can be used at any frequency where it is possible to construct the various components, and it has been found possible to construct mixers, amplifiers, harmonic generators and filters which operate up to several kilomegacycles per second. The maximum bandwidth over which they operate is limited by the regeneration of the harmonics at the bandpass filter.

There is a special case of this type of divider when $n = 2$. Then $(n-1)f = f$, and the harmonic generator and filter are no longer required as the ouput signal may be fed back into the mixer directly. Furthermore, it is often possible to combine the functions of the mixer and amplifier as in the circuit Fig. 7.[6] In this

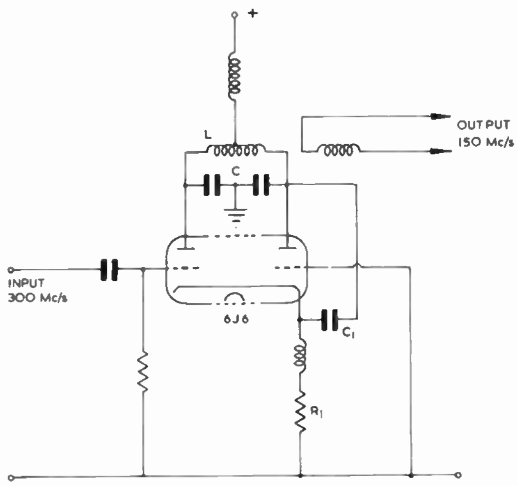


Fig. 7. Regenerative divider 300 Mc/s to 150 Mc/s.

circuit a 6J6 valve is used to divide by two a 300 Mc/s frequency. The circuit LC is tuned to 150 Mc/s and C is centred-tapped so that the signal fed to the cathodes through C1, R, will cause no resultant signal in the tuned circuit.

A 300 Mc/s signal applied at one grid causes the difference frequency to appear, in push-pull, at the anodes; and since the difference frequency is also 150 Mc/s, it establishes a feedback loop at half the input frequency. R1 is a wire wound resistor, and is shown as having both inductance and resistance.

2.3. *Counters*

Counters are used when the range of frequency to be divided is wide. The fastest counters in operation today employ mesa transistors, but it is expected that the application of tunnel diodes may lead to even higher speeds.

The circuit of a counter[7] with a resolving time of 5×10^{-9} seconds, i.e. which will count 200 Mc/s, is shown in Fig. 8. This is a modification of the circuit published by Chaplin and Owen.[8] VT1 and VT2 are the basic bistable circuit, and the transformer is the necessary element which ensures that the circuit will change state on being triggered. The high speed is obtained through the use of the avalanche diode VI to prevent VT1 from saturating and an extra transistor, VT3, to trigger the circuit.

The transistors in this circuit have an f_α of 600 Mc/s, and it should be possible to make faster counters as faster transistors become available. However, transistors may be soon replaced by Esaki or tunnel diodes in very high speed circuits. A tunnel diode counter circuit is shown in Fig. 9(a). The two diodes V1 and V2 operate as a bistable pair, and the resistors R1 and R2 provide a load line similar to that shown in Fig. 9(b). When the supply voltage is applied one diode will assume point A on the load line and the other point B. A positive pulse at the input will cause the diode at point A to switch to point C. The inductance L will ensure that the diode that was at point B will switch to the low voltage state and eventually reach

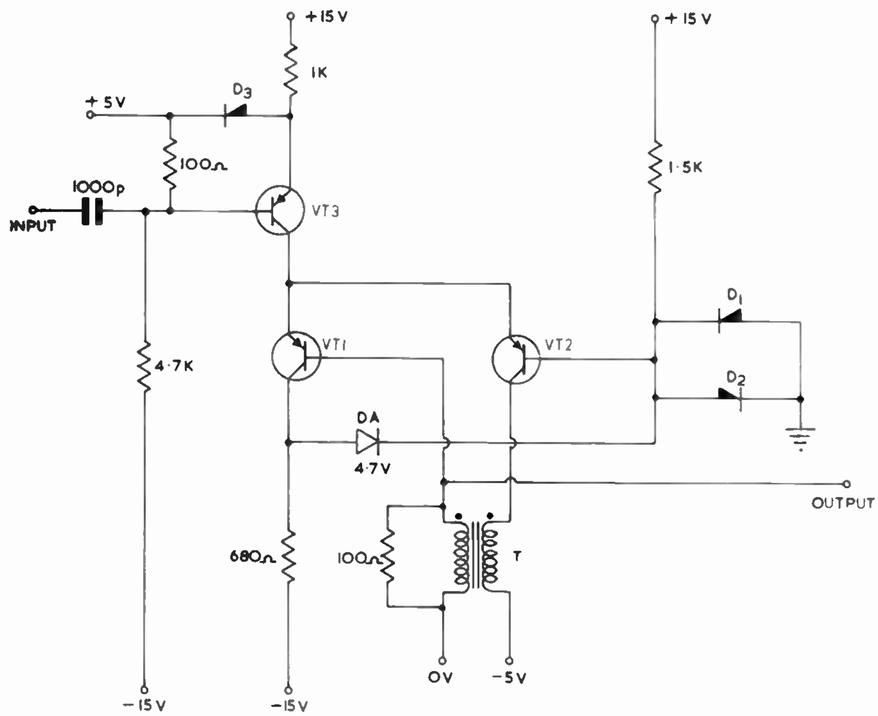


Fig. 8. High speed transistor counter.

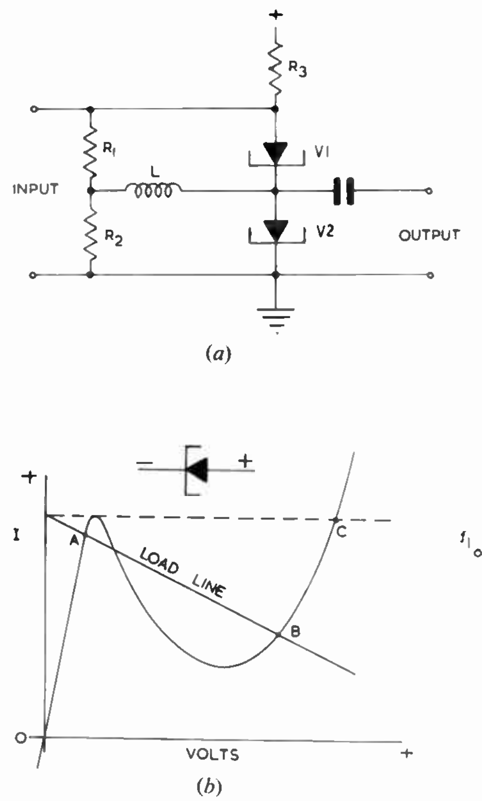


Fig. 9. (a) Tunnel diode counter circuit.
(b) Characteristic and load line tunnel diode.

point A, while the diode that was in the low voltage state will reach point B. The next positive pulse will cause the diodes to reverse their states again. If the load line is correctly placed the counter will not be switched by small negative pulses and one simple counter of this type will drive another.

When counters are used in chains to count large numbers, the resolving time of a single stage is often not the most important consideration in determining the maximum input frequency. The time taken for a carry to propagate down the chain, or the time taken to reset the counter at the end of counting a number, are often important. Frequently speed is lost in groups of four binary stages arranged to count to ten. Solutions to these problems are often found in a re-arrangement of the counter logic.[9] However, it is never possible to achieve the resolution of a single binary stage.

3. Mixers and Selectors

Additive and subtractive processes are carried out by mixers which function by modulation techniques. In frequency generation we are interested in the purity of the output signal. However, mixers produce spurious modulation products other than the major sum and difference ones generated in a balanced system. In fact harmonics of the input frequencies and the sum and difference of these harmonics in all combinations will appear in the output signal. A given output can be selected by tuning but the amount of attenuation of the other products is limited because with certain combinations of input frequencies the spurious products may be nearly coincident with the frequency of the desired output signal.

If we apply signals of frequencies f_2 and f_1 to an ideal linear multiplicative balanced mixer system then

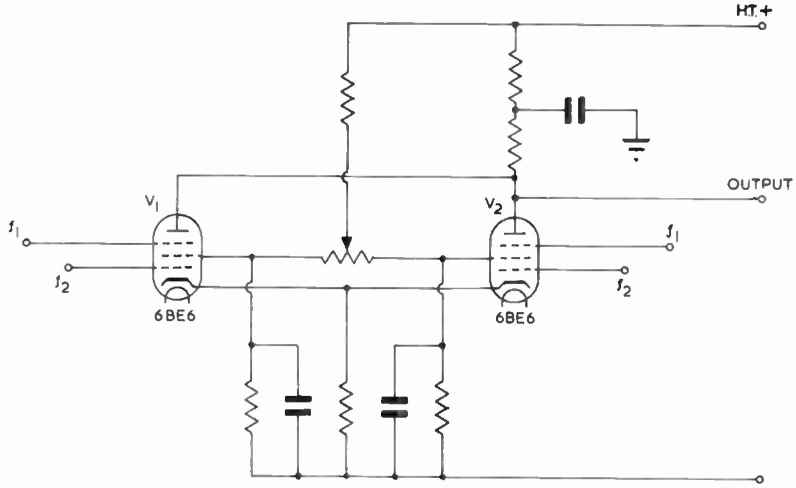


Fig. 10. Balanced linear multiplication system.

the output signals will be $f_2 \pm f_1$ each of equal strength. Figure 10 shows such a system where an approximately balanced and linear circuit has been realized. Table 1 gives the measured values of all the modulation products less than 80 dB of the level of the sum and difference products in such a circuit adjusted for

Table 1

Measured Attenuation Values of Modulation Products in Relation to $f_2 \pm f_1$ in an approximately Balanced System.

	f_2	30 dB	$2f_2$	30 dB	$3f_2$	33 dB	$4f_2$	58 dB
f_1 8 dB	f_2+f_1 f_2-f_1	0 dB 0 dB	$2f_2+f_1$ $2f_2-f_1$	37 dB	$3f_2+f_1$ $3f_2-f_1$	49 dB	$4f_2+f_1$ $4f_2-f_1$	68 dB
$2f_1$ 35 dB	f_2+2f_1 f_2-2f_1	38 dB	$2f_2+2f_1$ $2f_2-2f_1$	47 dB	$3f_2+2f_1$ $3f_2-2f_1$	69 dB	$4f_2+2f_1$ $4f_2-2f_1$	72 dB
$3f_1$ 42 dB	f_2+3f_1 f_2-3f_1	63 dB 65 dB	$2f_2+3f_1$ $2f_2-3f_1$	68 dB 66 dB	$3f_2+3f_1$ $3f_2-3f_1$	73 dB 72 dB	$4f_2+3f_1$ $4f_2-3f_1$	79 dB 78 dB
$4f_1$ 57 dB	f_2+4f_1 f_2-4f_1	60 dB 58 dB	$2f_2+4f_1$ $2f_2-4f_1$	72 dB 70 dB	$3f_2+4f_1$ $3f_2-4f_1$	77 dB	$4f_2+4f_1$ $4f_2-4f_1$	78 dB
$5f_1$ 71 dB								

the best operating conditions of bias and input signal levels. The table shows that it is necessary to consider fourth-order products of f_2 and f_1 to achieve an output $f_2 \pm f_1$ with all other products attenuated by at least 70 dB. Thus there are eight harmonics and thirty-two products which must be considered and the problem is to choose input frequencies not only to select $f_2 - f_1$ with respect to $f_2 + f_1$ or vice versa but to ensure that filtering of the spurious modulation products is possible, i.e. the products are not nearly coincident with the wanted one. When f_2 is $> f_1$ we can determine the forbidden ratios of f_2/f_1 to achieve the purity as given above and these ratios turn out to be 5/4, 4/3, 3/2, 5/3, 2, 5/2, 3, 4 and 5.

As the requirement for purity of the output becomes more stringent it is difficult to choose input frequencies such that high-order difference harmonics do not appear in the output. In the above treatment both input frequencies are considered as being pure sine waves.

This in itself limits the range of the frequencies which can be selected from a harmonic source derived from a frequency standard because the stability of the tuned circuits limits the number which can be cascaded to discriminate between high-order adjacent harmonics.

One method of selecting high-order harmonics has been suggested by Wadley[10] and used in communication equipment. Figure 11 shows a block diagram of the systems which uses two mixing stages with a drift-cancelled oscillator as the key element.

In this particular circuit the 15th, 16th, 17th or 18th harmonics can be selected. The first mixer is fed from the harmonic generator and from a substantially pure oscillator signal. Provided the oscillator is set close to the appropriate harmonic frequency then the output from the narrow band pass filter will be approximately a 6 Mc/s sine wave. Hence the signals appearing at the point of the second mixer will both be sine waves.

The oscillator frequency is used twice in this arrangement and its accuracy does not affect the final result with the proviso that its frequency must be sufficiently close to pass the narrow band fixed filter centred on 6 Mc/s. In the first mixing stage Δf, the frequency by which the oscillator deviates from the necessary harmonic setting, also appears in the 6 Mc/s output. However, in the second mixer the Δf cancels out leaving a 6 Mc/s component of the precise frequency of the harmonic generator. In the synthesis additive scheme (Fig. 1) it is only practical to select harmonics following the harmonic generator up to about the 10th order but the drift cancelled or catalytic oscillator enables a range of harmonics to be selected with a fixed filter selection circuit.

3.1. *Comparator and Control Systems for Slave Oscillator*

In order to control a slave oscillator a means of detecting its frequency error from a required frequency must be provided together with a control circuit for continuously correcting the existing error. The error frequency can be derived from a comparison with a sub-multiple of the slave's frequency and a reference frequency derived from the frequency standard as shown in Figs. 2 and 3. Camfield[11] has described a system for a 2000 channels communication equipment where a slave oscillator is motor driven to sweep through its frequency range until a phase discriminator and reactance valve take control to lock the slave's frequency precisely in terms of the frequency standard.

Phase discriminator and reactance valve circuits have been adequately treated in the literature. In general, control signals may be derived from either frequency or phase comparison circuits or preferably a combination of both. Frequency comparators involve the heterodyning of the two signals to obtain a difference frequency which can be rectified to drive a motor control of the tuning circuits to approxi-

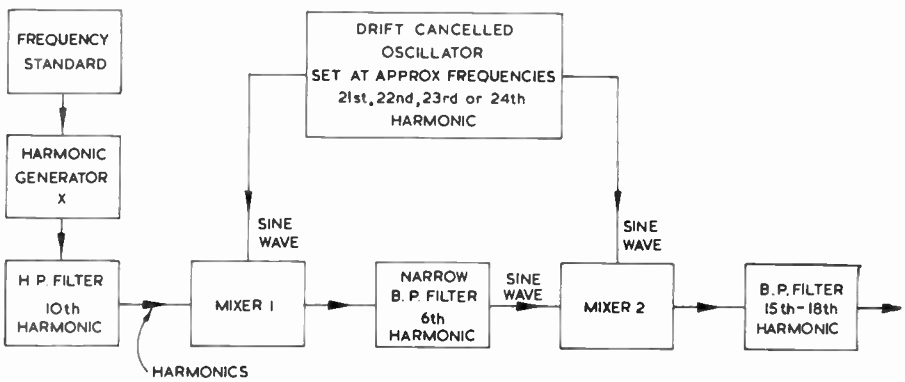


Fig. 11. High order harmonic selection by drift-cancelled oscillator.

mately the correct frequency. In this way it is possible to ensure that the reactance valve has a minimum pulling function. The phase comparison techniques added to the frequency control will enable the slave to be locked precisely but it should be remembered that there is always some time constant associated in the control loop. Consequently it is desirable to set the free resonant frequency of the slave as close as possible to the desired frequency.

If we assume that the counting or sampling technique on an indirect system is used for the spectrum below 1 Mc/s and that a synthesis or direct system is used above this frequency then the problem of frequency jitter must be considered. As a practical example the most significant decade control for the under 1 Mc/s frequency spectrum can be used to set L and C tappings of the slave oscillator to set its frequency to within 100 kc/s or 10% at 1 Mc/s. Given a time-constant of 1 sec for the control loop the frequency of the slave oscillator will perform about the 1 Mc/s frequency a sawtooth of peak amplitude of $\frac{100\,000}{1000}$ or 100 cycles providing we have a correction applied every millisec. From Fig. 12 it can be seen that the integrated effect of the shaded portion is $1/2000 \times 50/2 = 4{\cdot}5$ deg. phase shift, which is the maximum phase shift ever accumulated in the worst case.

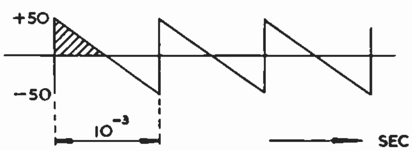


Fig. 12. Sawtooth error curve.

It is seen that, whilst this jitter only amounts to a very small phase error, its magnitude is controlled by the error in setting the natural frequency of the slave oscillator. A slow and simple motor-driven control loop can now be added to set the natural frequency closer to the desired figure.

The error waveform in the main loop takes the form of a sawtooth as shown in Fig. 12 and its polarity depends on whether the natural frequency is above or below the desired frequency. If, therefore, this waveform is combined in a phase sensitive rectifier with a suitable 1000 c/s waveform we can obtain a control signal to drive a "natural frequency-setting" motor. Supposing this extra loop sets itself to an accuracy of 5% of its maximum signal input (a fairly crude servo system) we can reduce the jitter assumed above by a factor of 20 times giving 2·5 cycles per second peak deviation or a maximum error of 0·2 deg.

4. Comparison of Direct or Indirect Techniques

In a direct system such as the additive or pure synthesis technique the frequency accuracy of any output generated is the same as that of the frequency standard. However, it has been shown that, because mixers produce spurious modulation products, it is not possible to generate signals of absolute purity. In practice these products will be of too high a level for a communication equipment except for single channel or where narrow frequency spectrums meet the communication requirements.

The indirect system on the other hand permits a signal to be generated which is substantially pure because it is possible to filter out spurious frequencies in the control loop. Moreover, the slave variable oscillator can be of higher power and the indirect system will need less amplification at the output frequency. The control loop time constant may impose some limitation on the output frequency but in general communication equipment is concerned with cycles per second rather than absolute accuracy. Because indirect systems also have the advantage of less equipment for a given frequency spectrum they are likely to be increasingly used for military equipments. At the present stage there is a good case for using indirect systems because of their simplicity even if the highest frequency end is carried out by synthesis. Developments of high speed integral counters are in hand for the division technique, but there is still a need for a simple equipment.

5. A Waveform Sampling System

A new and simpler method of frequency generation for communication equipment has been suggested by Oxford.[12] The method depends on time sampling the waveform of a slave oscillator and a simplified block diagram of the system is shown in Fig. 13. A frequency standard of 5 Mc/s is followed by the usual division chain to provide short pulses of 1000 per second. These pulses operate a clamp on the slave waveform which produces a control signal that is fed back to the slave oscillator. This closed loop system will set itself to produce a near zero voltage at the sampling intervals to give an integral number of cycles occurring in every millisecond. Hence when operating at any one of the desired channels 1 kc/s apart the slave oscillator has a repetitive waveform between sampling instants one millisecond apart. During each millisecond interval information can be acquired by further sampling processes which completely defines the slave frequency.

Near the start of the interval this information is approximate but unambiguous, and later it becomes more accurate although ambiguous. A combination of sampling points can be chosen which collectively

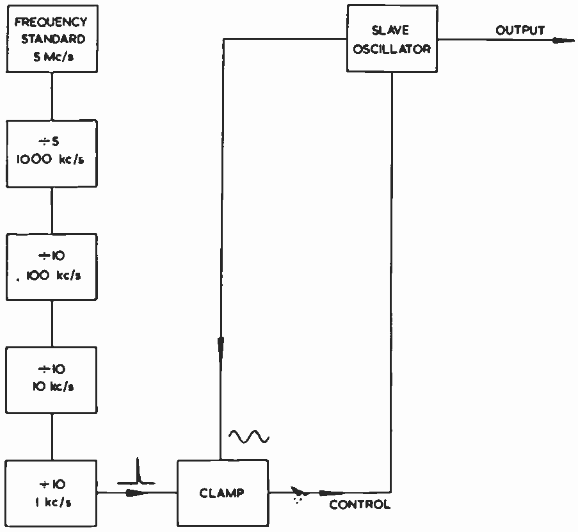


Fig. 13. Waveform sampling system.

give a complete description of the frequency, and the information obtained from each sampling point can be associated with the various stages of a decade control system in the following manner:—

Suppose we require to set the slave oscillator to 734 kc/s.

During each millisecond period it will perform 734 complete cycles. If, however, it is also sampled after only 100 microsec it will have done seventy three cycles and the voltage remaining on it at the sampling instant will be controlled solely by the extra 0·4 of a cycle. Similarly after a 10 microsec interval it will have performed seven complete cycles and the voltage at the sampling instant will depend only on the 0·34 of a cycle remainder. After 1 microsec delay the voltage sampled corresponds to 0·734 of a cycle.

These desired voltages can be set up on decade type potentiometers and by comparing them with the actual sampled voltages a control signal can be derived to correct the frequency of the slave oscillator.

6. Conclusion

Most methods surveyed in this paper have been proposed with the h.f. and v.h.f. bands in mind. This is the area where the problem is most pressing and where it has been recognized for some time past. For these bands combinations of straight synthesis with division or sampling techniques would appear to promise the most reliable and economical solutions, and developments in faster and faster switching elements will allow a gradual extension of the frequency band covered. At the same time there will be increasing demand for such equipment to be compatible in size, weight and power consumption with the smaller examples of communication equipment, although conforming to increasingly rigid specifications. It is unfortunate that the user is not generally prepared to operate in the binary system as this would in some cases result in apparatus economy.

It is not difficult to foresee that designers will be required to provide similar channelling equipment to control systems operating on frequencies up to several thousands of megacycles. At some stage it may become uneconomical to multiply crystal derived frequencies and special forms of cavities may be employed as frequency standards. New ideas would then be required in order to produce simple, flexible and reliable apparatus in the u.h.f. and s.h.f. bands.

As long as the demand for channel capacity continues to exceed the available spectrum there will be a need for channel allocating apparatus. It becomes more and more an integral part of any communication system and it is therefore right that communication engineers should find in it a fit subject for study.

7. References

1. R. T. Cox and E. W. Pappenfus, "A suggestion for spectrum conservation", *Proc. Inst. Radio Engrs*, **44**, pp. 1685–8, December 1956.
2. H. J. Finden, "Developments in frequency synthesis", *Electronic Engineering*, **25**, pp. 178–83, May 1953.
3. H. J. Finden, "The frequency synthesizer", *J. Instn Elect. Engrs*, **90**, part III, pp. 165–80, December 1943.
4. F. Stertzer and W. R. Bacon, "Parametric subharmonic oscillator", Paper 7–5, Solid State Conference, February 1959.
5. R. L. Miller, "Fractional-frequency generators utilizing regenerative modulation", *Proc. Inst. Radio Engrs*, **27**, pp. 446–57, July 1939.
6. P. M. Thompson, "Regenerative Divider 2/1", Roke Manor Report, December 1959.
7. A. R. Owens and A. J. Cole, "A High Speed Counter", A.E.R.E. Report, 1960.
8. G. B. B. Chaplin and A. R. Owen, "A junction-transistor scaling circuit with a 2 microsec resolution", *Proc. Instn Elect. Engrs*, **103B**, No. 10, pp. 510–15, July 1956.
9. R. W. Stuart, "A high speed digital frequency divider of arbitrary scales", *Conv. Rec. Inst. Radio Engrs*, Part 10, 1954.
10. T. L. Wadley, "Variable-frequency crystal-controlled receivers and generators", *Trans. South African Instn Elect. Engrs*, **45**, pp. 77–90. February 1954.
11. G. J. Camfield, "A frequency generating system for V.H.F. communication equipment", *Proc. Instn Elect. Engrs*, **101**, Part III, pp. 85–90, March 1954.
12. A. J. Oxford, "A waveform sampling system", British Patent No. 14167/60, March 1960.

Manuscript received 30th April 1960 (Paper No. 606.)

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. Information on translating services will be found in the Institution publication "Library Services and Technical Information".

"HUMAN" ENGINEERING

A recent issue of the *Journal* published by the Electronics Research and Development Establishment, Bangalore, contains two papers first presented at a Conference on Military Psychology held at Landour in June 1960. The first paper deals with the growth of complex modern systems and the role of the engineer in securing optimum performance from man-machine activity. The need for a study of "human" problems in engineering is examined, e.g. in the context of military "job" characteristics.

In the second paper it is pointed out that because of recent developments in technology and the need for higher speeds, an operator today is required to handle, operate and control a large number of diverse equipments and instruments which, if not properly preconceived and designed, will be beyond his capacity. The papers show an interesting approach to a modern problem.

"Human engineering", A. P. Bhateja (pp. 113–16), and "Human factors in equipment design", K. P. Singh (pp. 117–27). *Electro-Technology (Bangalore)*, 4, May-June 1960.

DETERMINATION OF TEMPERATURE COEFFICIENT

A simple, accurate laboratory test apparatus has been developed by the Department of National Defence, Ottawa, for measuring the temperature coefficient of electronic components in order to determine their suitability for use in stable tuning circuits. The paper announcing the device discusses the theory of the frequency shift method used and gives a full description of the equipment and its operation. It is claimed that an accuracy of 0·7 parts in 10^6 per deg C can easily be achieved.

"Simple apparatus measures temperature coefficient of components with high accuracy", C. Rempel and H. Reiche. *Canadian Electronics Engineering*, 4, pp. 40–43, 47, November 1960.

TRANSISTOR DECADE COUNTER

The design of a decade counter and of the associated diplay circuitry can be simplified through the use of multistable circuits. In the decade described in a French paper a bistable circuit has been combined with a circuit using a 5-state Lewis cycle. This arrangement yields a decimal division and offers some practical advantages over usual decades. The circuit is fully transistorized.

"Transistorized decade counting using the Lewis cycle", P. Balaskovic. *Electronique et Automatisme*, No. 5, pp. 227–30, September-October 1960.

TUNNEL DIODE CIRCUITS

The power gain and noise figure of two different circuits types of amplifier employing Esaki tunnel diodes are calculated in general terms in a recent German paper. It is shown that for the non-reciprocal (single stage) amplifier a noise minimum exists if the circuit is suitably proportioned. The result is extended to cover multistage amplifiers and this shows that a noise figure of about 5 dB can be realized for high gain amplifiers with one or several stages.

"Minimum noise figure for negative resistance amplifiers using Esaki diodes", M. Muller. *Archiv der Elektrischen Ubertragung*, 14, pp. 499–502, November 1960.

A COMBINED BACKWARD-WAVE OSCILLATOR AND TRAVELLING-WAVE TUBE

An electronically coupled combination comprising a backward-wave oscillator and a travelling-wave tube operating in the region of saturation is described in a German paper. Measurement results from such oscillators operating in the frequency bands around 4000 and 6000 Mc/s are reported. The efficiency values obtained correspond with those for travelling-wave amplifier tubes and are substantially greater than the efficiency of simple backward-wave oscillators. The output power level can be kept nearly constant over a wide frequency band by a suitable choice of the operating point of the travelling-wave amplifier system.

"An electrically tuneable microwave oscillator with a high efficiency and an output level independent of frequency", W. Eichin and H. Heynisch. *Nachrichtentechnische Zeitschrift*, 13, pp. 457–61, October 1960.

H.F. AERIALS

In a paper given at the 1959 Convention of the Australian I.R.E. a description was given of two types of aerial designed for launching vertically polarized waves in the 5–15 Mc/s frequency band over the sea at low angles. The aerials require a mast not higher than about 150 ft and are respectively of the triangular V and corner reflector types. It is claimed that the aerials, which are for use by the Department of Civil Aviation, are capable of gains exceeding 14 dB over the band with good matching and give efficient radiation between 1 and 10 deg elevation.

"Launching over the sea of vertically polarized waves for long distance ionospheric propagation", E. O. Willoughby. *Proceedings of the Institution of Radio Engineers Australia*, 21, pp. 591–7, September 1960.