

Founded 1925

Incorporated  
by Royal Charter 1961

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

VOLUME 41 No. 9

SEPTEMBER 1971

# THE RADIO AND ELECTRONIC ENGINEER

The Journal of the Institution of Electronic and Radio Engineers

## Automation and Control and the Electronic Engineer

**T**HE very word 'automation' conjures up widely different mental images to people in all walks of life. The social aspects are no doubt complex and far reaching, whatever the definition; however, to members of our Institution a narrow description—if not definition—may be acceptable, namely Control Systems, usually of a complex nature, almost inevitably highly electronic, and often using computers as system components.

The important thing about such a Control System is that it controls not itself but something else—a part of our external world, be it plant, aircraft, instruments, traffic and so on. The mantle of control system engineer often falls on the electronic engineer because his equipment can do more complicated things than anyone else's—particularly with the aid of a computer program, so he therefore must have a working knowledge as well of the ability and limitations of the power of stored program control. But the elegance of the electronics is as nought if the system does not effect the 'control' in a useful, effective and economic manner. It follows, therefore, that a knowledge of 'that controlled' is essential to some degree, and hence the control system engineer must know about many things other than the electronics of the control system. The other things he needs to know must include the transducers that connect his all-electronic control system to the outside world. Often the system cannot be all electronic and the control system engineer must be very aware of mechanical, electrical, hydraulic, optical and many other problems and solutions.

If automation depends heavily on complex control systems then the electronic engineer has a great responsibility both to himself, if he is to be an effective control system engineer, and to the public at large if the fruits of automation are to be harvested. For those whose interest or job lies in the control field an 'interdisciplinary knowledge' is thus essential. There are, in a sense, two levels of knowledge; the 'close' one of mechanical, electrical and other system component technology, and the infinitely wider one of the possibly-controllable outside world of traffic, chemical plant, aviation, radar, education and so on; the list is endless and growing!

Inevitably the wider knowledge is more dependent on the job of the engineer and the particular area of application he is concerned with. The Institution has always had close links with the medical world for example. The more 'narrow' knowledge of system components is of wider validity to a large number of individuals, and many in the field are endeavouring to arrange interdisciplinary events between the members of the institutions concerned. The I.E.R.E. is a member of long standing of U.K.A.C. (United Kingdom Automation Council), an association which has no 'corporate entity' other than that of its members who include most of the professional engineering institutions, accountancy and management bodies, as well as the Trades Union Congress and the Confederation of British Industry. Within this rather wide area of interest sub-groupings of member institutions can be and are possible, but the success of such groupings depends on the members of the institutions concerned. Greater activity will make for more, better, and hopefully, non-overlapping meetings. These are worthwhile not only for the formal papers presented but often are more valuable as a chance to gather and discuss common interests afterwards.

Current electronic technology is extremely competent and powerful, so that now the most important thing is to apply this technology and experience to improve the control of man over his environment. Much can be done if known techniques in one field can be put together in other fields to obtain new solutions. It is an exciting period despite temporary economic worries. The electronic engineer is uniquely placed to play an important role, but knowledge of electronics alone is not enough: the engineer must seize every opportunity to broaden his horizons.

A. ST JOHNSTON

## Contributors to this issue



**Professor G. R. Hoffman** graduated from St. Andrews University in 1950 with a B.Sc. degree in physics. In the following year he was appointed a member of academic staff at Manchester University and he has carried out research and teaching in the Electrical Engineering Department up to the present. He was appointed to a chair in electrical engineering in 1966. Professor Hoffman has contributed many

papers to learned and technical journals including *The Radio and Electronic Engineer*; most of these have been concerned with computer storage and display techniques.



**Mr. J. K. Birtwistle** was a development engineer with Ferranti Limited from 1951 until 1959 when he left to study electrical engineering at Manchester University. He graduated with a B.Sc. degree in 1962 and he has since then obtained his M.Sc. He has worked as a senior experimental officer in the Department of Electronic and Electrical Engineering at Manchester University since 1962.



**Mr. R. J. Royle** joined the Scientific Staff at the Post Office Research Station, Dollis Hill, in 1957 and worked on long term reliability of components until 1962. During this period he studied Applied Physics at Northampton Polytechnic, London, and Telecommunications at Willesden Technical College and Northern Polytechnic, gaining an H.N.C. in Applied Physics and a City and Guilds Full Techno-

logical Certificate in Telecommunications.

Granted a Post Office Sandwich Course Award in 1962, he subsequently gained an honours degree in Electrical Engineering at City University, London, returning to the Post Office Research Department in 1966 to take up a post in the Visual Telecommunications Branch where he is presently engaged in the design of experimental variable standard television equipment and projects allied to Viewphone research and development.



**Mr. C. N. O'Loughlin** joined the English Electric Valve Company as a junior engineer in 1955 on graduating from the National University of Ireland, Dublin. After working as a production engineer on transmitting tubes for five years, he transferred to the Power Beam Tube Department of the Microwave Division as a development engineer, being appointed assistant manager of the department in 1968 and manager in 1970.



**Dr. C. J. Edcombe** read engineering at St. Catherine's College, Cambridge, after gaining an English Electric Scholarship as an apprentice at the Marconi Company. Following graduation, he carried out research on a microwave device with the sponsorship of English Electric Valve Company and proceeded to the Ph.D. degree on 1966. Since then he has worked at English Electric Valve Company on high-

power microwave amplifiers, being concerned with their development and their behaviour in television systems.



**Mr. D. R. Bowman** (Member 1969) obtained his technical training as an apprentice with the U.K.E.A. at Windscale in Cumberland. He has held electronic engineering posts at Guy's Hospital, the D.W.S. station on Perim Island in the Middle East, the Radio Chemical Centre of U.K.A.E.A. where he was concerned with cyclotron particle accelerators, and as a senior engineer at the Aero-Space Division

of G.E.C. participating in solid-state microwave development work. He now holds the appointment of engineer-in-charge of the Electronics Section of the Chemistry Department at University College London.



**Mr. R. D. B. Waymark** (Graduate 1967) graduated in 1966 from the University of Bristol with the degree of B.Sc. in chemistry with electronics as a subsidiary subject. He also studied electronics as a day-release student at Enfield College of Technology during 1966-67, while working for Hilger and Watts Ltd. (now Rank Precision Industries) on electron spin resonance. Mr. Waymark joined the technical staff of the

Chemistry Department at University College London in 1969, where his main interest is nuclear magnetic resonance.

See also pages 397 and 420.

# Gravitational Wave Astronomy: An Interim Survey

By

**Professor P. B. FELLGETT,**

M.A., Ph.D., F.R.S.E.†

and

**D. W. SCIAMA,**

M.A., Ph.D.‡

The concept of gravitational radiation is introduced and the experiments of Weber are described. The magnitude and nature of the phenomenon are discussed in the light of Einstein's theory of relativity. The basic theory of a detector is developed and the means of realization are indicated. Investigations are proceeding along these and parallel lines in many parts of the world but no definitive results have yet been reported.

## 1. Introduction

Astronomy began before the dawn of history with visual observations of stars and planets. About a century ago, the range of effective astronomical observations began to be extended into the infra-red and ultra-violet regions of the electromagnetic spectrum. The first excursion into the radio-frequency spectrum was reported by Jansky<sup>1</sup> in 1932 in a paper modestly entitled 'Directional studies of atmospherics at high frequencies'. By 1935 the accumulated evidence enabled Jansky<sup>2</sup> to use the much bolder title 'A note on the source of interstellar interference'.

The thinking of radio engineers at this epoch was literally earthbound, and Jansky's results aroused surprisingly little interest, even when confirmed and extended by Reber between 1940 and 1944. Then in 1946 the warlike radars were beaten into the ploughshares of radio telescopes by Hey, van de Hulst, Pawsey, Ryle and others, and Radio Astronomy in the modern sense was born.

In all this development, our knowledge of the universe outside our own tiny planet has come entirely from electromagnetic radiation, with the sole exceptions of the Sun and Moon. If the Earth's atmosphere were opaque to the whole electromagnetic spectrum, or if mankind lived underground, the Sun and Moon could still be detected by their tidal effects. Indeed, no instruments are needed for this purpose, as every sea-bather knows.

If an analogy is drawn between gravitation and electromagnetism, tides correspond roughly to induction fields. Gravitation is however believed to be subject to the general rule that nothing capable of conveying intelligence can propagate at more than the speed of light, and accordingly it is theoretically possible for gravitational waves to exist.

Weber<sup>3, 4, 5</sup> has recently claimed to have detected an intense flux of just such gravitational radiation arriving at the Earth from a preferred direction in space. The analogy with Jansky's early observations is evident, and in this more sophisticated age intense interest has been aroused in the possibility that his results mark the beginning of astronomical observations using a quite new form of radiation to supplement what has been

learnt in the long history of observations of electromagnetic radiation. This interest extends to the whole question of gravitational radiation, its astronomical significance, and in improved techniques of detection.

The greater the novelty of a supposed observation, the greater the burden of proof it has to discharge, and at the present time Weber's results have yet to be confirmed by independent experiments. Detectors similar to Weber's, or believed to have better sensitivity, are currently being built in a number of laboratories, and this work has been characterized by fruitful and generous international collaboration.

Anything that can be said now must necessarily be tentative until the results of these new experiments are known. Weber's observations are however sufficiently suggestive to have justly aroused interest that makes it desirable for the whole subject to be widely discussed. In the remainder of the present paper, a brief account is given of the main questions that are involved.

## 2. Weber's Experiments

Weber's detector consists of an aluminium cylinder suspended in a vacuum chamber. Typical dimensions of the cylinder are 100 × 60 cm. Its lowest normal mode of longitudinal oscillation is 1660 cycles per second (hertz). This normal mode is excited both by thermal noise and by external disturbances, including the effects of possible gravitational waves. The presence of oscillations is established by means of piezoelectric detectors glued to the surface of the cylinder. These detectors convert their mechanical oscillations into electrical ones which can be electronically amplified and then recorded. Room temperature noise produces oscillations of the bar whose typical amplitude is only about 10<sup>-14</sup> cm, but Weber has succeeded in recording such tiny displacements which are ten times smaller than the radius of an electron.

In order to establish that a particular onset of oscillation is due to the passage of a gravitational wave, two things must be done. First of all it must be shown that the oscillation is due to an external signal and not to internally generated noise. This Weber achieves by having a second detector and looking for coincidences in the response of the two detectors. Secondly it must be shown that the signal is not due to seismic, electromagnetic, or cosmic ray disturbances. Subsidiary experiments can be performed to decide this question, for instance by monitoring the equipment, changing its electromagnetic coupling to the outside world, and so on.

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

‡ Department of Astrophysics, University Observatory, South Parks Road, Oxford, OX1 3RQ.

By means of detectors at the University of Maryland, and one at the Argonne Laboratory a thousand kilometres away, Weber has claimed to detect many events triggered by gravitational waves. At the time of writing he has been observing for over two years, and he claims an average of about one gravitational event per day. This exceeds by a factor of about 10,000 the rate of supernova explosions in the Galaxy. The events are not distributed uniformly in time as the Earth rotates, but cluster in the manner expected if the sources of the waves were at the galactic centre and the detectors have the directionality expected on Einstein's theory.

### 3. Gravitational Radiation

Einstein's theory of gravitation, which is the one generally accepted today, makes quite specific predictions about the properties of gravitational radiation. In many respects these properties are similar to the corresponding ones of electromagnetic radiation, although there are also some interesting and important differences. We shall therefore begin by recalling, extremely briefly, some of the essential properties of electromagnetic radiation. These properties are governed by Maxwell's equations (except for some quantum aspects which we mention later). In particular, electromagnetic waves in vacuum are transverse, and have two independent modes of polarization (for instance, left- and right-handed circular polarization). The propagation of a wave can be described in terms of rays lying at right angles to the wave front; these rays are straight lines in vacuum, and the wave can be regarded as moving along them with the fundamental velocity  $c$ .

According to quantum theory the rays can be regarded as made up of photons whose individual energy  $E$  is related to the frequency  $\nu$  of the waves by Planck's relation  $E = h\nu$ . Each photon moves with the fundamental velocity and may be regarded as spinning around an axis lying along its direction of motion, the two directions of spin corresponding to the two directions of circular polarization of the wave in the Maxwell picture. The angular momentum associated with the spin of each photon is unity in the conventional unit  $\hbar = h/2\pi$ .

We can also think of electromagnetic waves in relation to their sources, namely charged particles which are accelerating relative to an inertial frame. In addition to its Coulomb field, an accelerating charge produces a field that decreases inversely as the first power of the distance. The associated flux of energy is determined by the square of this field, and so decreases by an inverse square law. The total flux of energy across a sphere is thus independent of its radius, and we may speak of energy being radiated to infinity (although in practice much of the energy would be absorbed by other charges on the way).

As a final preparation for our discussion of gravitational waves, it is helpful to look at electromagnetic waves from yet another point of view. We can regard them as the means by which the *information* is conveyed to other charges that the distribution of source charges has changed in some way, so that their total Coulomb field has also changed. Such a change in the Coulomb

field would not be felt by other charges instantaneously, but only after a delay corresponding to propagation of the information at the speed of light. This so-called retardation effect is a deep property of special relativity. It is important to note that the change in the source could not be a change in its total charge; this would violate the law of conservation of charge which is an important feature of Maxwell's theory. The simplest possible change in the source is one in which the *moment* of the charge distribution about a fixed point changes. A change in this so-called dipole moment is permitted by Maxwell's theory, and would lead to the emission of (dipole) radiation. In more complicated cases the second or higher moments of the charge distribution in the source can change, and the angular pattern of the radiated intensity is correspondingly more complicated.

We now turn our attention to gravitational waves, where considerations of this type are of great importance. Einstein's field equations differ from Maxwell's in two ways:

- (i) there are ten potentials instead of four (the electric potential  $\phi$  and the three components of the magnetic vector potential  $\mathbf{A}$ ), i.e. 'tensor' theory instead of vector theory;
- (ii) the gravitational field acts as one of its own sources, so that Einstein's equations are non-linear.

When the gravitational field is weak it is a good approximation to neglect (ii) and we then arrive at Einstein's weak field equations. These equations are the immediate generalization of Maxwell's equations from four potentials to ten. Accordingly they imply that changes in the gravitational field are propagated as transverse waves moving with the speed of light. The quantization of this weak-field theory is straightforward (by contrast to the problem of quantizing the full non-linear theory), with the result that gravitational rays may be regarded as being made up of gravitons obeying the relation  $E = h\nu$ . Each graviton spins either right-handedly or left-handedly around its direction of motion, but because of the extra number of potentials it turns out that the angular momentum of each graviton is 2 in units of  $\hbar/2\pi$ .

Einstein's weak-field equations also enable us to relate the emission of gravitational radiation to the acceleration of a source particle. Indeed a small accelerating mass produces a gravitational field which decreases inversely as the first-power of the distance. As in the electromagnetic case we expect this to lead to the radiation of energy essentially to infinity. The only difference is in the angular distribution of the radiation, which is that associated with a tensor theory.

Finally, we can regard gravitational radiation in terms of the propagation of information about changes in the source. This point has been put with particular force by Bondi.<sup>6,7</sup> Suppose that the Sun suddenly ceased to exist. How long would it be before the orbit of the Earth was affected? The answer, of course, is that the Sun cannot cease to exist. Mass, like charge, is conserved. But we can go further, for a theory with ten potentials contains more conservation laws than a theory with four potentials. In fact, the mechanical analogue of the

electric dipole moment is conserved. The mechanical moment, that is, the moment of the mass-distribution, is in fact zero if it is taken about the mass-centre (that is how the mass-centre is defined, of course). In the absence of external forces the mass-centre cannot suddenly accelerate; it remains permanently at rest in a particular inertial frame because of the extra conservation law, the conservation of momentum. The mechanical (dipole) moment thus remains zero in an inertial frame and so there can be no dipole radiation in the gravitational case. In terms of the Sun as a source we have therefore answered the question: what would happen to the Earth's orbit if the Sun suddenly moved off at right angles to the orbit? The answer is that the Sun cannot do that because of the conservation of momentum.

What the Sun can do is to change its shape. The *second* (or quadrupole) moment of the mass-distribution need not be conserved, so that as a result of internal forces the Sun could suddenly become a spheroid. If it did so, it would then emit the simplest type of gravitational radiation, namely quadrupole radiation. The Earth's orbit in the field of a spheroid is, of course, different from its orbit in the field of a sphere, and the difference would begin not instantaneously, as in Newtonian theory, but a light-time later. In fact the Earth's orbit would begin to change at the same instant that we would *see* the Sun change its shape. From the relativistic point of view this is a very satisfactory result.

Our discussion so far has been entirely qualitative. We now turn to the quantitative question: how intense do we expect gravitational waves to be in practice? Our first expectation is that they should be extremely *weak* in comparison with the electromagnetic waves that occur in practice. The reason is that gravitational interactions are very much weaker than electromagnetic ones. For example, the ratio of the gravitational to the electrical force between an electron and a proton,  $Gm_e m_p / e^2$ , is about  $10^{-39}$ . (The gravitational force of the Earth far exceeds its electrical force only because the positive and negative charges in the Earth tend to cancel each other out, whereas their gravitational forces act in the same direction.) Consider now the gravitational waves emitted by the solar system. The planets in their motion round the Sun are expected to emit gravitational waves. Most of the power is produced by Jupiter, and it amounts to about 450 watts. A similar power is produced by the atoms of the Sun, which accelerate in each other's electrical fields. By contrast, the power of the electromagnetic waves emitted by the Sun is about  $10^{24}$  times greater.

More intense gravitational waves are probably emitted by a catastrophically collapsing mass, perhaps associated with a supernova explosion. In such cases most of the power is emitted when the object has collapsed into the strongly non-Newtonian regime near its Schwarzschild radius  $2GM/c^2$  (although not so close that the gravitational red shift becomes large enough to rob the waves of most of their energy). Accordingly we would expect to receive at the Earth a pulse whose duration is at most a few times longer than the time taken for light to cross

the Schwarzschild radius of the collapsing object. This latter time is close to  $10^{-5} M/M_\odot$  seconds. Thus a collapsing star of ten solar masses would produce a gravitational pulse lasting perhaps less than a millisecond.

It is very difficult to calculate the expected energy in such a pulse. The reason is that if the star remains spherical no gravitational waves at all are emitted as a result of the collapse (there would be no change in quadrupole moment and in fact the gravitational field outside the star would remain exactly the Schwarzschild field with a constant mass). In practice the star is likely to be non-spherical because of the effects of rotation and possibly internal magnetic fields. The resulting asymmetric collapse is very difficult to follow in general relativity, and we do not know what fraction of the star's mass is radiated away in the form of gravitational waves. A plausible estimate is a few percent, which would imply for our ten solar mass case that nearly one solar mass equivalent of energy may be radiated in gravitational waves. The true efficiency may be even higher in some cases.

We must now consider how these waves might be detected. This is the same question as considering how energy can be extracted from them, so that they become at least partially absorbed. This question is more complicated than in the electromagnetic case because we must allow for the effects of the Principle of Equivalence. This Principle implies that a uniform gravitational field must induce the same acceleration in all parts of a proposed detector of gravitational waves, and in the observer himself. In principle the observer could watch the distant stars with great precision in order to observe their acceleration relative to him as a wave passes over him. In practice this acceleration would be too small to detect. It is therefore necessary to look for *relative* accelerations between different parts of the detector, and these must be caused by *non-uniformities* in the gravitational field of the wave. In a similar way, ocean tides on the Earth are raised by non-uniformities in the Newtonian gravitational field of the Moon and the Sun, and so we may speak of *tidal* forces acting on the detector arising from any non-uniformities in the gravitational field of the wave. Fortunately, non-uniformities will be present in the wave, and their tidal effect leads to relative motions in the detector which could, for instance, generate heat. This argument shows clearly that energy is carried by gravitational waves and can be extracted from them and converted into more familiar types of energy. It also shows how the waves can be detected by relatively localized observations.

What this argument does not indicate, however, is how small a fraction of the gravitational wave energy falling on a detector would be absorbed by it. We expect this fraction to be small for the same reason that the gravitational power emitted by a non-catastrophic source is small, namely, the relative weakness of the gravitational interaction. Consider, for example, the whole Earth acting as a detector of incoming gravitational waves. Dyson<sup>8</sup> has calculated that for waves with a frequency of 1 Hz the Earth absorbs only about  $10^{-21}$  of the gravitational energy falling on it. This shows clearly

that matter is extremely 'transparent' to gravitational waves, and that it is correspondingly difficult to build a practical detector.

#### 4. Gravitational Waves in the Non-linear Theory

Our considerations so far have been based on the equations obtained when the self-interaction of the gravitational field is neglected. These equations are linear and easy to handle. The neglected terms are usually extremely small, the most important exception being the case of catastrophic gravitational collapse. Nevertheless, it is of great theoretical interest to study the behaviour of gravitational waves when the non-linearities are taken into account, and much recent effort has gone into the study of this problem. We shall describe a few of the highlights here.

No exact solution of Einstein's field equations is known which represents gravitational waves being emitted from a material system. However, as a wave recedes from its source a limited portion of the wavefront approximates more and more closely to that of a plane wave. An exact solution for such a wave was found independently in 1956-7 by Robinson and by Bondi. The discovery of this solution caused quite a stir at the time. Very few exact solutions of any kind were then known, even fewer had any physical significance, and the status of gravitational radiation in the full theory was quite uncertain. It was even claimed by some that the radiation phenomenon did not occur at all in the full theory. The discovery of an exact solution representing radiation, even in the unrealistic case of a plane wave, showed quite rigorously that gravitational waves do exist in general relativity.

Nevertheless the plane wave solution is misleading in some respects. The wave propagates without distortion and also without dispersion. If, for instance, one makes a 'sandwich' wave in which the disturbance is initially confined between two plane wavefronts moving in the same direction with the speed of light then the disturbance retains its shape indefinitely and never leaks out from the sandwich, remaining confined between the two plane wavefronts. This and other properties of the plane wave solution were discussed in a basic paper by Bondi, Pirani and Robinson<sup>9</sup> in 1959. The lack of dispersion is particularly misleading in that it is a special property which is unique to plane waves. More general waves develop a tail behind the wavefront. These tails have very complicated properties which make exact calculations too difficult to perform.

Nonetheless, a good deal is now known about the properties of non-plane waves, but only when they are far from their sources. Indeed, this asymptotic theory is in many respects complete. Much less is known about the relation of the waves to their sources. The key result here is due to Bondi, van der Burg and Metzner,<sup>10</sup> who showed in 1962 that if a material system is initially static, changes its shape, and becomes static again, then it loses mass. The mass-loss is proportional to the square of the rate of change of the field, summed through the period of change. Note that this is an essentially *non-linear* result. Energy is radiated in gravitational waves

resulting from the change of *quadrupole* moment, and this energy loss shows up as a reduction in the total *mass* of the source. In a linear theory the different moments of the mass distribution would behave quite independently; here the change in the second moment leads to a change in the zero-th moment (the total mass). Note also that the mass-loss does not *cause* the radiation, but is caused by it. It is the change of shape that causes the radiation.

Finally we should mention the results of a recent non-linear study by Thorne<sup>11</sup> of the non-spherical oscillations of a star. These oscillations lead to the emission of energy in the form of gravitational radiation. As would be expected, the energy in the oscillations is found to decrease at the same rate as energy is being radiated (unless of course the oscillations are further damped by ordinary dissipative processes within the star). Considerations of this sort now play an important part in the interpretation of that most remarkable of astrophysical phenomena, the pulsars.

From the theoretical point of view, therefore, there is no difficulty in principle in accepting Weber's results. There is, however, a difficulty of practice. For Weber claims to observe about one significant event per day, and this leads to two astronomical difficulties. The first is that the rate of events is unexpectedly high, and so we have to fall back on the supposition that processes occurring close to the galactic centre are quite different from those we can study in detail in nearer parts of the Galaxy. The second is that the Galaxy must be losing energy and mass by gravitational radiation at an extraordinarily large rate, of the general order of several hundred solar masses per year, that is, at a rate about  $10^4$  times the total optical power of the Galaxy in all starlight. Neither of these difficulties is insurmountable, but astronomers would not be willing to make the necessary drastic changes in their outlook until Weber's claims are tested by independent experiments of adequate sensitivity. To this experimental problem we now turn.

#### 5. Detection of Gravitational Radiation

A detector of electromagnetic radiation has a set of positive and negative charges, crudely the electrons and atomic nuclei of the antenna, on which the incident radiation exerts a differential force. There is no gravitational analogue of charges having opposite sign, and a putative gravitational force at any point in space is unobservable by the Principle of Equivalence. An antenna for gravitational radiation must therefore depend on detecting the relative acceleration of two masses  $m, m'$  separated in space by a distance  $l$  (Fig. 1).

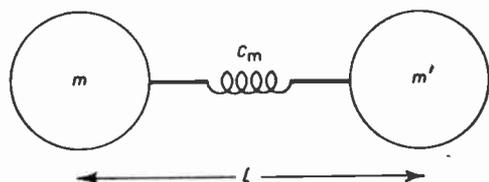


Fig. 1. General arrangement of gravitational detector.

The observed relative force is

$$F = m \frac{d^2 l}{dt^2} = mc^2 l R_{0101}$$

where  $R_{0101}$  is the appropriate component of the Riemann tensor, and  $c$  the velocity of light. In practice there will be some compliance  $C_m$  connecting the two masses. The sensitivity varies with direction of arrival according to a  $\cos 2\theta$  law as shown in Fig. 2.

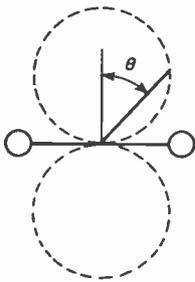


Fig. 2. Directivity of gravitational detector.

It is next necessary to transduce the force  $F$  to an electrical signal. Hitherto this has always been done piezoelectrically, although a number of other possibilities exist, including capacitive and optical methods, and these ought to be explored. The currently preferred piezoelectric configuration (Gibbons and Hawking<sup>12</sup>) has the general form of a cylindrical bar split equatorially with the piezoelectric transducer sandwiched between the two halves (Fig. 3).

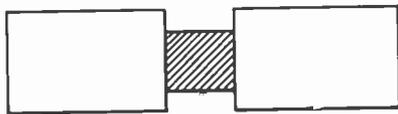


Fig. 3. Configuration of piezoelectric gravitational antenna. (The piezoelectric material is shown shaded.)

An approximate equivalent circuit for this arrangement is shown in Fig. 4. It uses the formal analogy between force and voltage, and between velocity and current. The electrical and mechanical circuits are linked by a notional transformer the 'turns ratio' of which is not dimensionally a pure number. In this representation the values assigned to the mechanical components are those directly measurable mechanically and do not depend directly on the responsivity of the transducer. In Fig. 4  $r_{eq}$  is an equivalent resistance

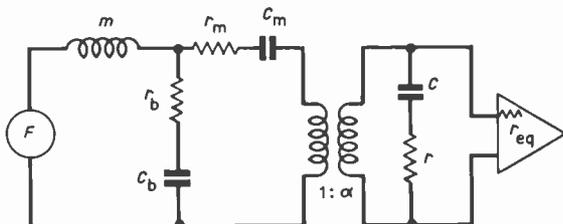


Fig. 4. Approximate equivalent circuit of gravitational antenna.

representing the amplifier noise voltage,  $C$  the electrical capacitance of the transducer and  $r$  the loss associated with it,  $C_m$  the mechanical compliance of the transducer and  $r_m$  its associated loss, and  $C_b$  and  $r_b$  represent the compliance and loss of the bar itself. The notional transformer has a 'turns ratio'

$$\alpha = \frac{A d}{l s}$$

equal to the piezoelectric charge generated per unit displacement, where  $A$  is the area of cross-section of the transducer,  $s$  is modulus of compliance, and  $d$  is the piezoelectric strain constant (short-circuit charge per unit area produced by unit tensile stress). The coefficient  $d$  is the measure of piezoelectric activity most commonly quoted in the literature.

The optimization of the design can follow the standard procedure of first increasing the useful signal as much as possible relative to the sources of noise that are either fundamental or unavoidable in practice (for example, because materials of unlimited rigidity or conductivity do not exist), and then arranging that the noise output comes predominantly from only these noise sources. Evidently the signal  $F$  can be increased by increasing the mass  $m$  or the separation  $l$ , but a larger  $l$  may increase undesirably the shunt compliance  $C_b$  and its associated noise-generating loss  $r_b$ . The ability of the signal to override the noise generated by the electrical loss  $r$  in the transducer and the amplifier noise (represented by  $r_{eq}$ ) is influenced by its frequency in relation to the resonance of  $m$  with approximately  $C_m + C_b$ .

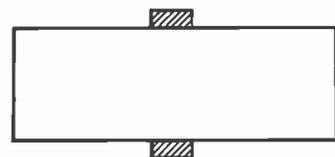


Fig. 5. Schematic layout of Weber gravitational detector.

In the Weber form of detector, the piezoelectric material is glued to the equator of an undivided bar, as shown in Fig. 5. This has the effect of capacitively tapping-down the signal source, as indicated in Fig. 6, and is less favourable.

Evidently it would be advantageous to resonate  $m$  within the band of frequencies that the gravitational signal is expected to occupy. This introduces a dilemma,

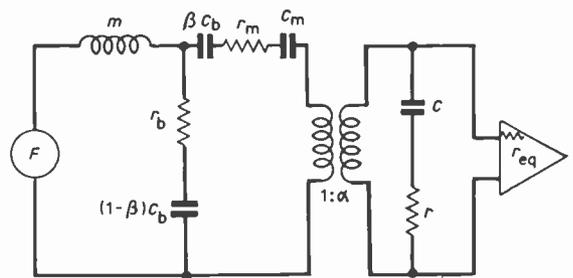


Fig. 6. Approximate equivalent circuit of Weber detector ( $\beta \approx 2l_T/l$  where  $l_T$  = length of transducer).

since the signal characteristics are inherently unknown until the signal has been unequivocally detected. Theoretically the most likely source of gravitational radiation is either the gravitational collapse of a body of stellar mass, or the capture of one collapsed object by another. The simplest waveforms consistent with these mechanisms of generation are sketched in Fig. 7. The radiation is expected to be emitted with a time scale  $\tau$  of the order of  $10^{-5} M/M_{\odot}$  seconds, where  $M$  is the mass of the object and  $M_{\odot}$  the mass of the Sun.

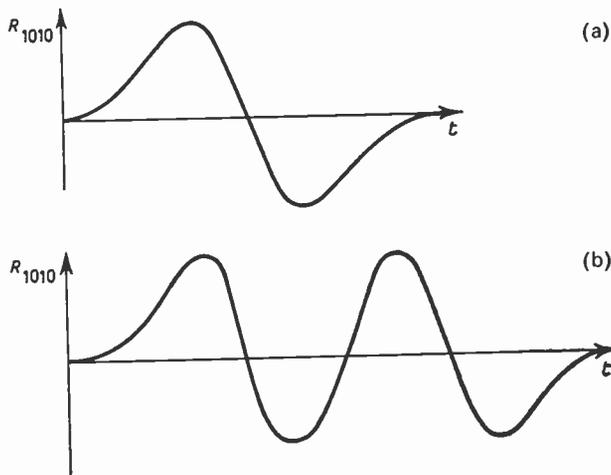


Fig. 7. Expected gravitational signal from (a) gravitational capture; (b) collapse.

If the sources are distributed fairly uniformly in our Galaxy, their average distance from the Earth is about the same as that of the centre of the Galaxy. Calculation shows that an energy of about  $6M_{\odot}$  would have to be radiated at this distance in order to produce a signal detectable by Weber's antenna. The mass of the source itself is then likely to be about  $100M_{\odot}$ , giving  $\tau \approx 1$  ms and a wideband signal peaking in the region of  $f = 1$  kHz. This is near to the frequency at which Weber and most other investigators have resonated their detectors. It is however difficult to believe that an object as massive as this collapses, or captures another, on average every day as Weber's observations seem to imply. Perhaps it would be better to make the detector most sensitive to the occasional slow event that might be produced by a super-massive object, or alternatively to concentrate on fast events produced by less massive objects galactically close to the Earth. At this stage the engineer may well feel that speculation has replaced solid design data, and that it is time to make empirical trials.

Analysis indicates that available piezoelectric materials (and amplifiers) enable the antenna of Figs. 5-6 to exhibit a good noise factor over a bandwidth  $B$  of the order of the resonant frequency  $f$ ;  $B \approx f \approx 1$  kHz. This gives a time resolution of around 1 ms, and the emission of  $1M_{\odot}$  at the galactic centre could then be detected with a false alarm rate of only once per annum.

Double or triple coincident detection of even one such event in widely spaced receivers would demonstrate the existence of gravitational radiation in the Galaxy with

virtual certainty. This performance implies the detection of an increment of energy of only  $\sim 1/250$  of the thermal energy  $kT$  of the detector mechanical system (Gibbons and Hawking<sup>12</sup>). This is made possible by increasing the  $Q$  of the system as much as material properties allow, so that the mechanical mode exchanges energy with the thermal field only slowly.

## 6. Evaluation and Prospects

At the present time, antennae designed to detect gravitational radiation are being built or are in operation in the U.K., U.S.A., U.S.S.R., Italy, Switzerland and probably other countries as well throughout the world. Theoretically, it would be very surprising if gravitational radiation did not exist and fulfil the essential criterion of Physics of being observable at least in principle. Nevertheless, at the time of writing Weber's observations still lack confirmation by independent observations of good signal/noise ratio. Moreover, a flux as large as claimed by Weber does indeed present formidable theoretical difficulties. The verdict at the present time must be 'not proven', but it is very probable that new experimental data will resolve the present uncertainties one way or the other within the next year.

Subsidiary evidence can be obtained from consideration of phenomena that are likely to accompany the violent gravitational events supposed to generate Weber's observed events. A co-operative search has been made for coincident radio-frequency events at 151 MHz in Europe (Charman *et al.*<sup>13</sup>), and an independent search in the U.S.A. at 19 GHz (Partridge<sup>14</sup>). These two searches have yielded negative results, and they respectively set upper limits of approximately  $10^{-14}$  and  $10^{-22}$  to the ratio of the radiation sought to that required to trigger Weber's detector. It is almost inconceivable that collapsing plasma should fail to emit radio radiation. Perhaps the sources are collapsed objects (black holes) with negligible atmospheres which capture one another. Although a number of other reasons can be adduced for the failure to observe coincident radio events, nevertheless these negative results appear puzzling.

## 7. Acknowledgments

The authors are greatly indebted to all their colleagues who have supplied information on which this interim account is based, and in particular to the members of the Science Research Council Panel to Co-ordinate Research into Gravitational Radiation.

## 8. References

1. Jansky, K. G., 'Directional studies of atmospherics at high frequencies', *Proc. Inst. Radio Engrs*, **20**, pp. 1920-32, December, 1932.
2. Jansky, K. G., 'A note on the source of interstellar interference', *Proc. Inst. Radio Engrs*, **23**, pp. 1158-63, October 1935.
3. Weber, J., 'Evidence for discovery of gravitational radiation', *Phys. Rev. Letters*, **22**, pp. 1320-4, 16th June 1969.
4. Weber, J., 'Gravitational radiation experiments', *Phys. Rev. Letters*, **24**, pp. 276-9, 9th February 1970.
5. Weber, J., 'Anisotropy and polarization in the gravitational radiation experiments', *Phys. Rev. Letters*, **25**, pp. 180-4, 20th July 1970.

6. Bondi, H., 'Gravitational waves', *Endeavour*, **79**, pp. 121-30, July 1961.
7. Bondi, H., 'Assumption and Myth in Physical Theory', p. 63 (Cambridge University Press, 1967).
8. Dyson, F. J., 'Seismic response of the earth to a gravitational wave in the 1-Hz band', *Astrophys. J.*, **156**, pp. 529-40, May 1969.
9. Bondi, H., Pirani, F. A. E. and Robinson, I., 'Gravitational waves in general relativity. III: exact plane waves', *Proc. Roy. Soc. A.*, **251**, pp. 517-33, 23rd June 1959.
10. Bondi, H., van der Burgh, M. G. J. and Metzner, A. W. K., 'Gravitational waves in general relativity. VII: waves from axi-symmetric isolated systems', *Proc. Roy. Soc. A.*, **269**, pp. 21-52, 21st August 1962.
11. Thorne, K. S., 'Nonradial pulsation of general-relativistic stellar models. III: analytic and numerical results from neutron stars', *Astrophys. J.*, **158**, pp. 1-16, October 1969.
12. Gibbons, G. W. and Hawking, S. W., 'The detection of short bursts of gravitational radiation', *Phys. Rev.* (to be published).
13. Charman, W. N., Fruin, J. H. and Jelley, J. V., Hodgson, E. R., Haynes, R. F., Scott, P. F. and Shakeshaft, J. R., Baird, G. A., Delaney, T. J. and Lawless, B. G., Drever, R. W. P. and Meikle, W. P. S., 'A Search for Isolated Radio Pulses from the Galactic Centre at 151.5 MHz'. Circulated S.R.C. report. May 1971 (PCRGR 31).
14. Partridge, R. B., 'Search for microwave pulses associated with gravitational radiation', *Phys. Rev. Letters*, **26**, pp. 912-15, 12th April 1971.

*Manuscript received by the Institution on 7th July 1971. (Paper No. 1404/IC48.)*

© The Institution of Electronic and Radio Engineers, 1971

## The Authors



Professor P. B. Fellgett took Natural Science Tripos at Cambridge University, and between 1948 and 1951 read for a research degree on infra-red and other measurements at the University Observatories. After a year at Lick Observatory of the University of California where he worked on photoelectric measurement of stellar magnitudes, he returned to Cambridge for a further seven years to work on

spectrometry and later on computer analysis of astronomical measurements. From 1959 to 1965 he was at the Royal Observatory, Edinburgh, latterly as Head of the Astronomical Instrumentation Division where he was responsible for the development of the 'Galaxy' automatic measuring engine for Schmidt photographs.

Appointed to the Chair of Cybernetics and Instrument Physics in the Department of Applied Physical Sciences at the University of Reading in 1965, Professor Fellgett has carried out research into classification and retrieval of information, transport problems and various aspects of instrument physics, including direct digital transduction of environmental properties which formed the basis of a paper he

presented at the Institution conference on Digital Methods of Measurement in 1969.

He holds a number of industrial consultancies and is a member of the Home Secretary's Committee on Prison Security; he has recently become Chairman of the Science Research Council Panel to Co-ordinate Research into Gravitational Radiation.



Dr. D. W. Sciama was educated at Trinity College, Cambridge, where he received a B.A., M.A. and Ph.D. and was awarded a Research Fellowship in 1952. Since then he has been a Visiting Professor at Cornell University, an Agassiz Fellow at Harvard and a member of the Institute for Advanced Study at Princeton. More recently he was appointed to a Lectureship in Mathematics at Cambridge and a Fellowship of

Peterhouse (from 1961-70). In October 1970 Dr. Sciama took up a Senior Research Fellowship at All Souls College, and he is also a member of the Department of Astrophysics at Oxford University, where he is building up a theoretical group in astrophysics, relativity and cosmology.

# A Method of Measuring Pulse Duration

By

R. J. ROYLE, B.Sc.(Eng.)†

The pulse under investigation and a timing wave consisting of a succession of narrow triangular pulses of controllable frequency are simultaneously displayed on a double-beam sampling oscilloscope. Adjacent timing pulses are aligned to the leading and trailing edges of the pulse being measured; the duration of the pulse, being then identical with the period of the timing wave, is obtained from a digital frequency meter monitoring the timing-wave frequency. Pulse durations in the range 20 to 100 ns may be determined with an uncertainty of  $\pm 100$  ps. For pulses of durations extending to several microseconds the uncertainty is within  $\pm 0.1\%$ . The method has particular application in pulse-bandwidth measurements on television and other waveform transmission systems having pulse bandwidths in the range 125 kHz to 25 MHz. Using ancillary amplitude-measuring equipment, such measurements can be accomplished with uncertainties of  $\pm 1\%$  or better.

## 1. Introduction

### 1.1 Origin of Investigation

In television and other waveform transmission systems, the concept of 'pulse-bandwidth' as proposed by Lewis<sup>1, 2</sup> offers several advantages over the more conventional definitions of bandwidth because it is a criterion by which primary effects of bandwidth limitation in a system can be judged.

In this concept pulse bandwidth is defined for lowpass systems as the reciprocal of twice the duration of the rectangular pulse whose response has a peak amplitude equal to 87.27% of the final amplitude of the step response.

This definition has merit in that it includes some practical degree of frequency weighting, takes account to some extent of phase distortion and has useful meaning when applied to markedly nonlinear systems; also, the particular amplitude ratio (more precisely  $(2/\pi) \text{Si}(\pi/2) = 0.872654$ ) has been chosen to equate the pulse bandwidth to the cutoff frequency in the special case of an ideal lowpass filter.

To exploit these and other advantages offered by the pulse-bandwidth concept and to assist in its wider acceptance as a practical definition, it is necessary that convenient and sensitive means exist for its measurement and in particular for the accurate determination of pulse duration inherent in such a measurement.

It was to fulfil this particular need that the work which is the subject of this paper was instigated. However, it is felt likely that the results of the work will prove of more general use in other applications where precise determinations of pulse durations are required.

### 1.2 Design Objectives

In conventional pulse duration measurements, where it is common to measure the pulse against a graticule on a real-time oscilloscope, inaccuracies can occur as a result of

- (i) parallax errors due to a finite distance between graticule and cathode-ray tube screen
- (ii) time-base imperfections such as non-linearity and inaccuracy of calibration.

(iii) resolution errors due to the necessity of displaying the pulse as a whole on a relatively narrow screen typically 10 cm wide.

Such errors can accumulate in practice to a total of one or two per cent in the final result of a pulse-duration measurement.

Since an accuracy preferably better than 1% is considered a desirable design objective for a complete pulse-bandwidth measuring system, an accuracy considerably better than 1% is required of the pulse duration part of the measurement, thus necessitating the quest for a more sophisticated form of measurement.

The method adopted here, which obviates the errors mentioned above is based on the production of narrow timing pulses of a frequency which is both controllable and continuously displayable on a digital frequency meter.

For the measurement, adjacent timing pulses are accurately aligned with the leading and trailing edges of the pulse under test on a wide-band double-beam sampling oscilloscope having time expansion and wide-range time-position facilities; the required duration of the pulse is then obtained from the frequency meter.

With this basis of measurement, it is necessary that the shape of individual timing pulses is sharply peaked and that the half-amplitude duration is preferably about 1 or 2 ns so that precise registration with the leading and trailing edges is possible when measuring pulses of duration as short as 20 ns, the desired range of measurement being 20 ns to about 10  $\mu$ s.

To facilitate speedy alignment, another design requirement is that one timing pulse should remain in automatic registration with the leading edge of the pulse being measured while the adjacent timing pulse is brought into coincidence with the other edge. The amplitude of the timing pulses should remain constant and independent of frequency for the same reason.

Automatic continuity of display as the frequency of the timing pulses is adjusted is a further objective of design, i.e. synchronism and stability of oscilloscope traces should be maintained throughout a measurement without recourse to external control.

The choice of the repetition rate of the trigger pulses, which have to be generated and applied to the pulse

† Post Office Telecommunications Headquarters, Research Department, Dollis Hill, London, NW2 7DT.

generator supplying the test pulse, is not critical but a value in the region of 16 kHz is suitable as it is not so low that it causes undue display flicker but is low enough to permit measurements of pulse duration extending to several microseconds. A trigger repetition rate of about 16 kHz also allows ample time interval between the test pulses when measuring pulse bandwidths in the 0.5 to 25 MHz range on systems having pulse responses with long-duration ringing. Such an interval prevents discernible overlap in the responses of the system to consecutive test pulses.

It is of course implicit in a technique involving the continuous synchronism of a pulse with a frequency-adjustable timing wave that some variation in the pulse repetition rate must occur and that only a pulse generator whose output pulse duration is independent of such variation is suitable for accurate measurements. Variation in trigger repetition rate has nevertheless to remain within controlled limits so that trigger controls on the oscilloscope need no adjustment during measurement. Derivation of the trigger pulses from the timing wave by a form of variable division, as opposed to a fixed division, is thus essential.

**2. Operation**

Figure 1 is a simplified schematic showing the equipment which evolved in accomplishing the design objectives.

The block diagram enclosed by the dashed rectangle indicates the principal functions of the circuit designed to produce the timing pulses and the 15.6 kHz trigger pulses. This portion of the design has been developed and constructed as a self-contained unit. The remaining part of the diagram external to the dotted contour represents additional proprietary equipment and interconnections involved in a typical pulse-duration measurement, the basic system for which is as follows.

The output of the variable-frequency oscillator, after intermediate conversion from sine to square waves is transformed by the timing-wave generator into the timing

pulses which are applied via a delay line to channel 1 of the sampling oscilloscope (Fig. 2(a)).

A variable-ratio divider fed by the timing-pulse generator produces a nominal 0.5 MHz train of pulses which after further division by a fixed-ratio divider triggers both the oscilloscope and test-pulse generator at a nominal 15.6 kHz rate; the timing pulses, being therefore synchronous with the test pulse applied to channel 2, may be aligned to the edges of the pulse as shown in Figs. 2(b), (c) and (d) by adjusting the oscillator frequency. Pulse duration is then obtainable from the frequency meter connected to the oscillator.

In the case of a pulse-bandwidth measurement, the width of the test pulse is set by first switching S to position 2 and applying a simulated step waveform to the network being measured in the form of a long-duration test pulse; this duration is then reduced to the setting at which the observed response falls to 87.27% of the final, or quasi steady-state, value attained by the step response. The switch S is returned to position 1 and the oscillator frequency adjusted to align the markers with the test pulse. After the alignment, the pulse bandwidth is read directly from the meter measuring one-half the oscillator frequency.

**2.1 Square-wave Generator**

The circuit of the square-wave generator is shown in Fig. 3. Sine to square-wave conversion is carried out by applying the sinusoidal output of the external oscillator to a parallel combination of two current-mode operated Schmitt-trigger stages. Parallel operation is achieved by combining collector and base leads of appropriate transistors and employing common input, output and Zener-diode coupling circuitry.

The design operates over a frequency range extending to above 50 MHz and provides an unloaded output of 3 V across 50 Ω with rise and fall times of 1.1 ns and 1.3 ns respectively.

The output voltage of the combined Schmitt-trigger stages is increased by an emitter-current switch which is

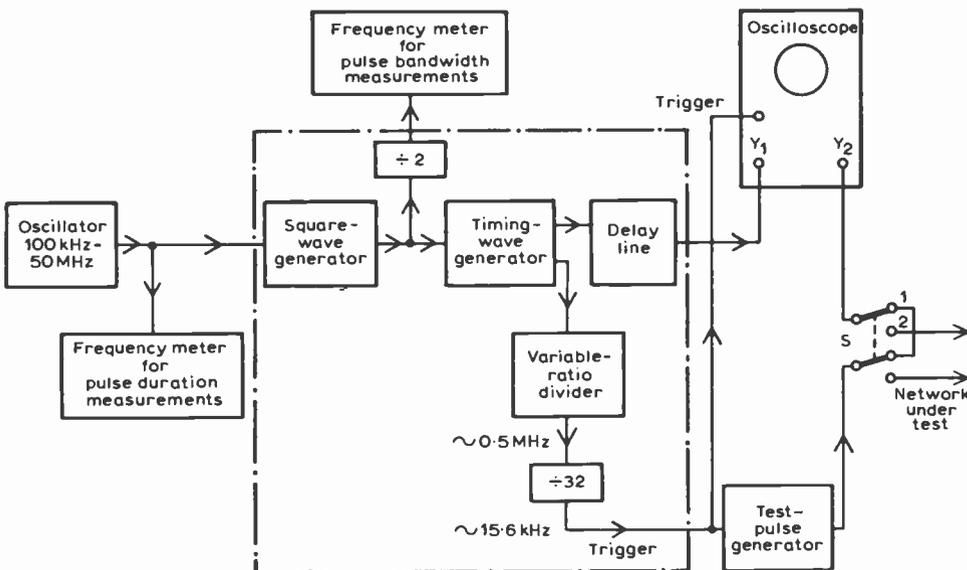
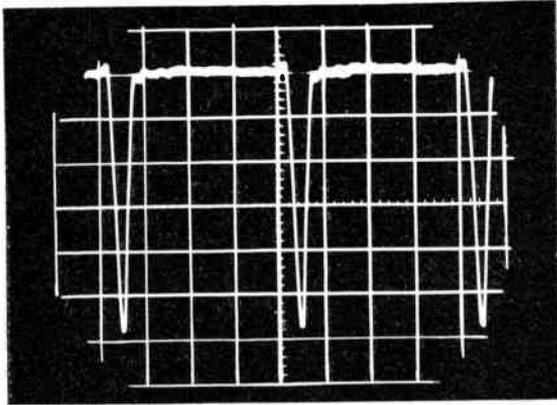
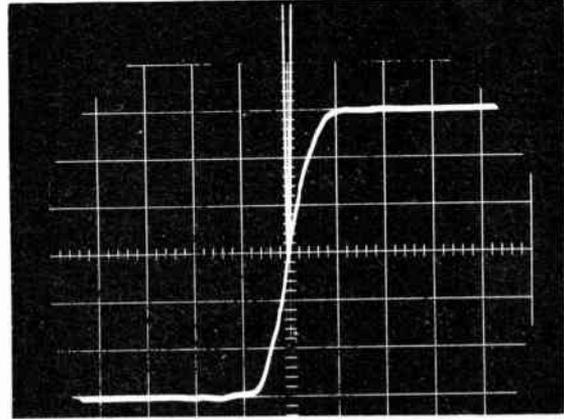


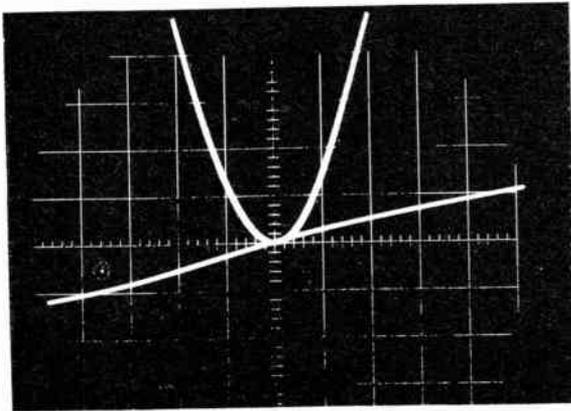
Fig. 1. Method of measuring pulse duration (simplified schematic).



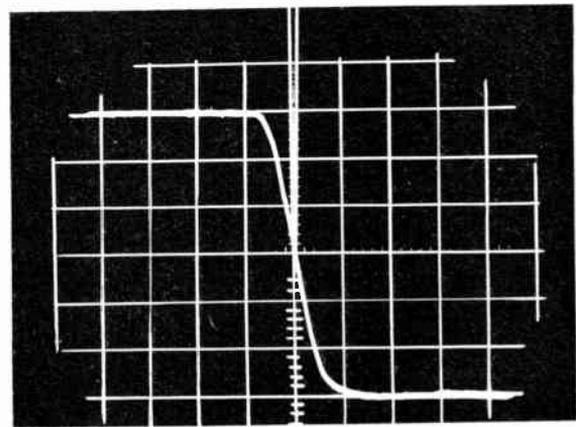
(a) Scale 200 mV/div. vertical, 5 ns/div horizontal.



(c) Scale 50 mV/div vertical, 5 ns/div horizontal aligned to leading edge of test pulse.



(b) Scale 50 mV/div vertical, 200 ps/div horizontal aligned to leading edge of test pulse.



(d) Scale 50 mV/div vertical, 5 ns/div horizontal aligned to trailing edge of test pulse.

Fig. 2. Timing pulses.

of conventional long-tailed pair type except that a parallel combination of two transistors replaces each of the transistors in the normal arrangement.

### 2.2 Divide-by-2 Circuit

Normally, for a pulse-duration determination, the digital frequency meter is driven direct from the sine-wave output of the external variable frequency oscillator, the reciprocal of the frequency display giving the pulse duration.

For convenience in pulse-bandwidth measurements an alternative drive for the meter is available from an emitter follower fed from a high speed bistable; this bistable divides by two the frequency of the square wave appearing at the current switch stage so that with this connexion, the numerical result of a pulse-bandwidth measurement is displayed directly on the meter.

An incidental advantage of this alternative connexion is that it may permit the use of a less costly frequency meter as the maximum operating frequency is reduced by a factor of two.

### 2.3 Timing-wave Generator

The output of the square-wave generator drives a two-diode pulse generator circuit in which two silicon switching diodes, of a type which have been discovered to have a reverse transient characteristic similar to

snap-off diodes, are utilised (with an appreciable cost advantage) in an adaptation of the circuit described by St. Marie.<sup>3</sup>

In the modified circuit shown in Fig. 4, D1 and D2 are both forward biased with D1 having slightly less bias. With the output of the square-wave generator applied to the input, diode D1 cuts off abruptly at a time after each positive input step which is determined by the exhaustion of its stored charge, thus causing the potential at the base of TR9 to rise sharply. Diode D2, having

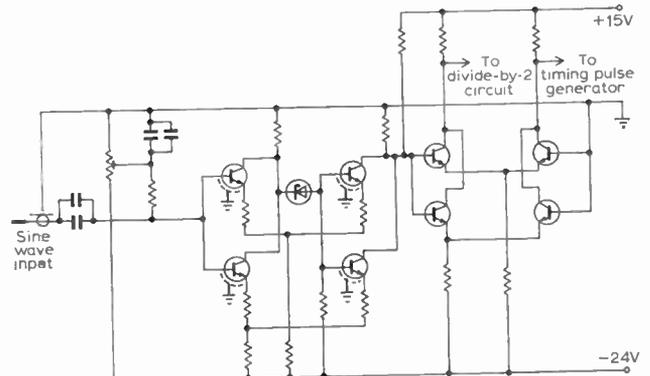


Fig. 3. Square-wave generator.

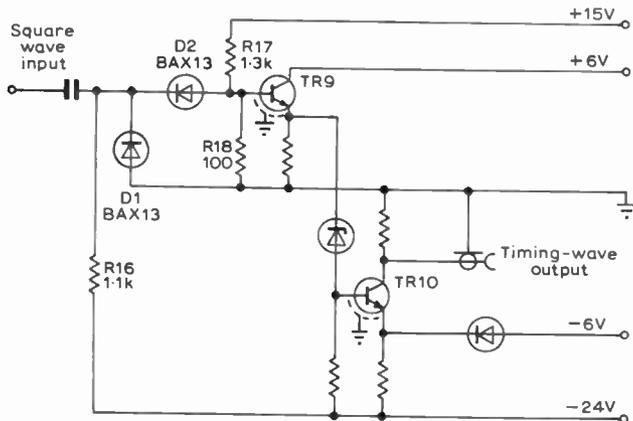


Fig. 4. Timing-wave generator.

slightly greater stored charge as a result of greater initial bias, cuts off correspondingly later causing the potential to return to near earth. As a consequence a succession of positive pulses is produced, the period being equal to that of the square wave.

Since pulse duration is determined by the difference in storage times of two identical diodes, the detrimental effect of a high temperature-coefficient of stored charge on pulse duration stability (a problem in some step-recovery diode circuits) is minimal.

Bias resistors R16 and R17 are carefully selected to make the difference in storage times so short that immediately diode D1 reaches cut-off, diode D2 starts to turn off. In this manner, very sharp pulses are produced across R18 which, without the loading of subsequent stages, have an effective height of 1.4 V and a half-amplitude duration of 1 ns.

Feeding this pulse train via a base-clipping emitter follower TR9 into a pulse height discriminator TR10 serves to remove spurious spikes from the pulse train and results in a timing wave of negative pulses on a substantially clean base-line, suitable for connexion by 50 Ω cable to one channel of the sampling oscilloscope (Fig. 2(a)). The length of this cable is carefully selected to have a specific propagation delay in order to satisfy the design requirement of maintaining one timing pulse in permanent alignment with the leading edge of the pulse being measured. This is explained more fully in Section 2.6.

2.4 Variable Ratio Divider

The technique used for the variable ratio divider is an adaption of a gating-method of division employed in the pulse-and-bar generator described by Macdiarmid and Phillips.<sup>4</sup>

An additional train of negative pulses of 1 ns duration similar to the timing pulses and likewise derived from the two-diode pulse generator is applied to input 1 of the current-mode two-input NAND gate shown in Fig. 5.

Unless inputs to the base of TR11 and TR12 are both at logical one (negative) level, the 'constant' current flowing through R24 is steered through one or both transistors and the common collector resistor R23 causing the collector potential to be low.

Interruption of transistor conduction occurs whenever a negative pulse at TR11 base is time coincident with a negative level at TR12 base and the momentary re-routing of current via hot-carrier diode D4 produces the positive output pulse at the collector junction.

In operation a negative level on input 2 enables the inverted passage of a single pulse through the gate. This transmitted pulse is used to initiate the closure of the gate to all subsequent pulses at input 1 for a nominal period of 2 μs, after which time the gate is re-opened to allow the next pulse arriving at input 1 to pass and instigate a further closure.

To accomplish this repetitive sequence of division, the output pulse from the gate is stretched by a tunnel-diode monostable to a pulse of approximately 30 ns duration and this pulse used to operate another monostable, this time of nominal 2 μs period and capable of a 100% duty cycle.

Current switches following the latter monostable reduce the rise and fall times of the monostable output pulse and apply to the gate, input 2, a positive-going pulse approximately 2 μs wide, with rise and fall times of 3 ns and 1 ns respectively which inhibits the gate for the 2 μs period; this period starts 20 ns after the time a transmitted pulse transverses the gate because of propagation delay round the loop.

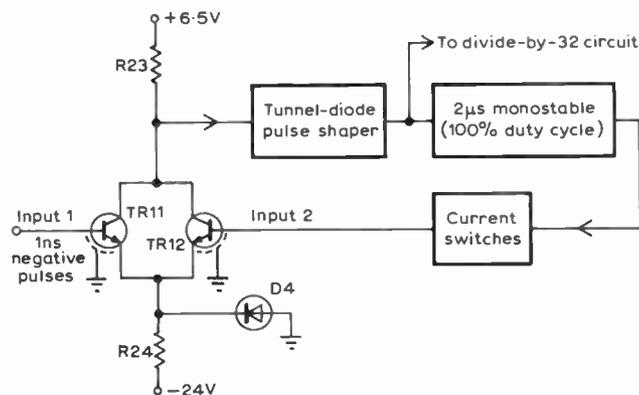


Fig. 5. Variable-ratio divider.

The successive 'gating out' of pulses for periods of 2 μs gives the result that a pulse train of repetition frequency  $f$  MHz (in the range 1–50 MHz) applied to input 1 of the gate produces at the output of the tunnel diode monostable a train of stretched pulses of repetition frequency  $f/n$  MHz,  $n$  being the smallest positive integer which makes  $f/n \leq 0.5$  MHz. If  $f$  is reduced gradually then  $f/n$  decreases proportionally until the next lower value of  $n$  satisfies the inequality. The output frequency then jumps to 0.5 MHz, the division ratio  $n$  having changed automatically to the next lower value. The maximum departure of the output frequency from the nominal 0.5 MHz when  $f$  is varied is  $-1/2(1+2f)$  MHz.

These stretched pulses are coupled to a 'divide-by-32' circuit whose output pulses, now with a nominal frequency of 15.6 kHz and a possible maximum excursion

of  $-1000/64(1+2f)$  kHz,† are used for triggering the test-pulse generator and the sampling oscilloscope.

### 2.4.1 Synchronism enhancement

For all variable-frequency 'divide-down' circuits which have a division ratio which increases in integral steps with increase of input frequency, it must be accepted that points of uncertainty in division, and hence instability in circuit operation, will exist in the transition from one division ratio to the next.

In the system, as described so far, the points of instability occur during adjustment of the timing pulse frequency when the trailing edge of the pulse used for re-opening the gate, applied at input 2, is time-coincident with a pulse appearing at input 1. The resulting uncertainty at these positions as to whether the pulse will pass or not is apparent on the oscilloscope as loss of synchronism.

No instability, of course, arises due to time-coincidence of a pulse at input 1, with the leading edge of the pulse at input 2 as this is prevented by the propagation delay between a pulse leaving the gate and the arrival of the leading edge at input 2 being no greater than 20 ns—the minimum period of the input pulse train.

The length of the uncertainty period that does arise and the consequent degree of non-synchronism noticeable is reduced by making the input pulses to the gate as narrow as possible to reduce the chance of their being 'cut' by the gate and by employing current switches

† To avoid this excursion becoming excessive when measurements of pulse duration in the microsecond region are attempted, two non-adjacent timing pulses may be used for alignment purposes and the period corresponding to the frequency-meter display multiplied by  $n+1$ , where  $n$  equals the number of timing pulses separating the aligned pulses.

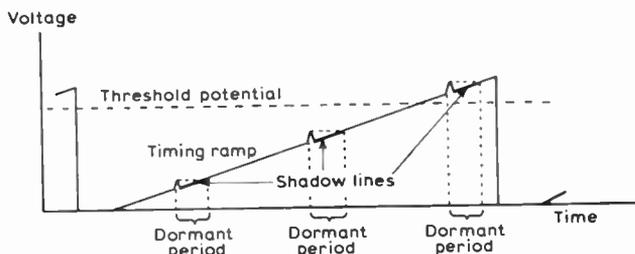


Fig. 6. 2 μs timing ramp, showing shadow lines.

following the 2 μs monostable to speed up the re-opening edge of the pulse and thus quicken the action of the gate.

Another palliative is, however, necessary because of instability, observed at certain critical settings of input frequency, which results from random fluctuations in the arrival time of the trailing edge of the pulse at gate input 2, caused by about 0.03% jitter in the 2 μs monostable period.

Successful treatment consists of adding to the voltage-ramp, which performs the timing function in the 2 μs monostable, a train of narrow pulses of suitable phase and amplitude derived from the timing-pulse train.

Since the nominal 2 μs period ends when the timing ramp reaches a pre-set threshold potential, one of these pulses may terminate the ramp if it occurs just before threshold is reached. If the pulse positions are such that they fail to do this, they produce 'shadow lines' on the ramp, i.e. a range of position following each pulse (Fig. 6), during which no termination of the monostable period is possible—a fact independent of threshold level or the relative pulse position on the ramp.

During a dormant period the initiation of the rear edge of the monostable 2 μs pulse is prevented and the consequent arrival of an inhibiting pulse at gate input 2 is prohibited for a similar length of time.

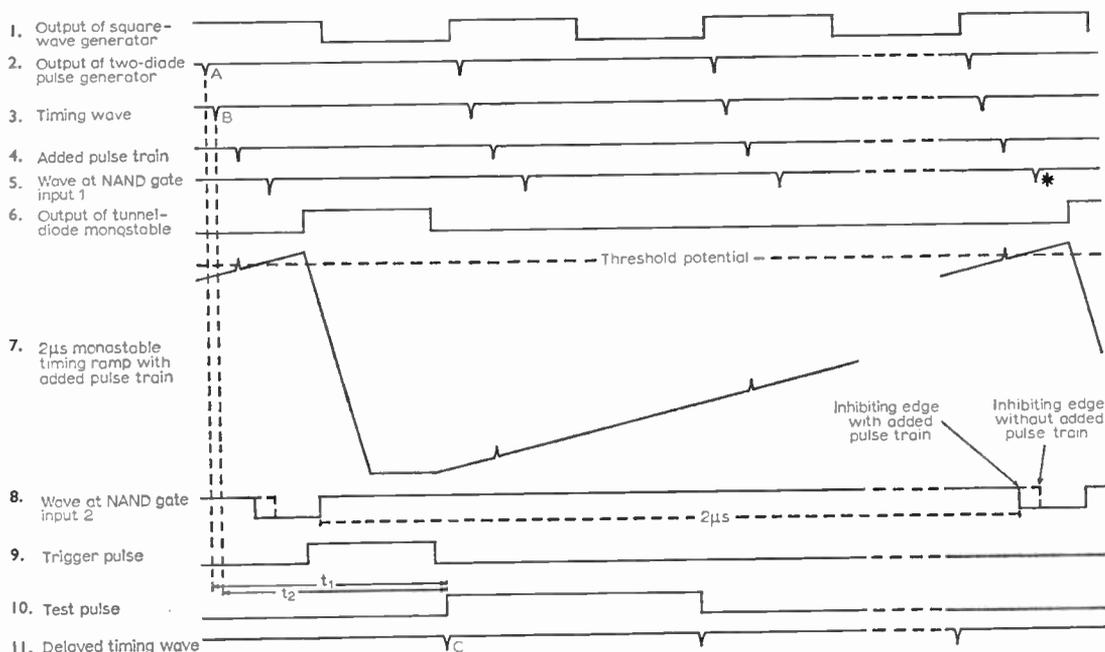


Fig. 7. Waveform diagram (not to scale).

Adoption of the figure 0.005 for the ratio of the amplitude of the added pulses to threshold potential makes the dormant time-span equal to 10 ns, and incorporation of a short delay cable prior to gate input 1 trims the timing of the pulse train at this input such that each pulse arriving there appears mid-way through the 10 ns period during which the presence of an inhibiting pulse edge at input 2 is prohibited.

As a consequence, random fluctuations in arrival times of the inhibiting pulse edge at input 2 of the gate only occur at non-critical times when no pulses are due for transmission; instability is thus not experienced.

The effectiveness of the 'shadow-line' technique is apparent in Fig. 7 which indicates the shape and relative timing (though not to scale and excluding minor propagation delays) of waveforms existing at various stages throughout the system.

Without the pulses shown added to the timing ramp (waveform 7) but with the same threshold potential and frequency setting, the inhibiting pulse edge of the waveform at gate input 2 (dotted edge, waveform 8) is time coincident with pulse (shown asterisked, waveform 5) at gate input 1. Consequently the 'gating-out' of this pulse tends to be haphazard being dependent on any noise, spurious pulse breakthrough or slight fluctuations in period.

With pulses added to the timing ramp, however, the inhibiting pulse edge is pre-triggered precisely, and occurs at a safe instant well in advance of the pulse to be transmitted.

#### 2.4.2 2 $\mu$ s monostable

The circuit diagram of the 2  $\mu$ s monostable showing the method of adding the pulses to the timing ramp is shown in Fig. 8.

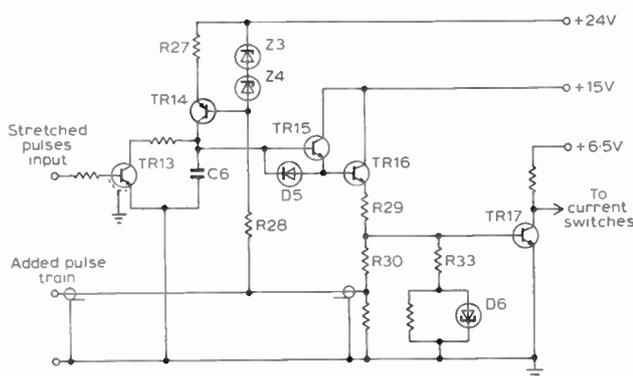


Fig. 8. 2  $\mu$ s monostable.

*Correction.* The negative end of R28 should be shown connected to the earth line, not to the 'added pulse train' line.

In the initial absence of positive pulses from the tunnel-diode pulse-shaper stage, transistor TR13 is off and timing capacitor C6 is charged almost linearly from a constant current generator formed by transistor TR14, Zener diodes Z3, 4 and resistors R27, 28.

The voltage ramp thus produced across C6 is transformed by the Darlington-pair<sup>5</sup>, TR15 and TR16, into

a sympathetically-increasing current in R29 which, in the early stages of growth, flows mainly through tunnel diode D6, this being in a low-voltage state. As the potential on C6 nears a limiting value of about 14V, the peak point value (10 mA) of the current in D6 is attained and an abrupt change from low to high voltage state takes place. Transistor TR17 in association with D6 forms a threshold detector and, being biased by the sum of the potential across D6 and that across R33, is turned on by the sudden change of state. In the continued absence of pulses from the pulse-shaper stage, a quiescent condition is reached with C6 fully charged and TR17 conducting strongly: this condition identifies the stable state of the monostable.

The stable state is terminated or prevented from being reached by the transmission of a positive 30 ns pulse from the pulse shaper to the base of TR13. This pulse serves to discharge timing capacitor C6 which remains discharged until the end of the 30 ns period, after which charging of C6 recommences. The discharge of C6 is communicated via the Darlington pair to the threshold detector causing D6 to return to the low voltage state and TR17 to cease conducting. Diode D5 improves the speed of this communication by providing a path for removal of base charge in TR16 during turn-off.

The sum of the input pulse duration (30 ns) and the time taken for C6 to recharge to the level corresponding to the critical potential of the threshold detector establishes the nominal monostable period.

The addition of pulses to the timing ramp is effected by summing, at the base of TR17, the current ramp passing through R29 and the diminished timing pulses passing through R30.

Since the input pulses to TR13 are themselves instrumental in discharging and resetting the timing capacitor to zero, the need for the recovery period normally required in monostable circuits is eliminated and a 100% duty-cycle made possible.

#### 2.5 Divide-by-32 Circuit

Five J-K bistables, using integrated circuits and driven from the nominal 500 kHz, 30 ns pulse train from the tunnel-diode pulse-shaper stage, are connected in tandem and have their individual outputs gated with the input in a six-diode AND gate. The resulting 15.6 kHz 30 ns pulse train at the output of the AND gate, after passage through a Darlington pair, is available for the required external triggering functions.

#### 2.6 Timing-wave Delay

Further examination of Fig. 7 shows that the train of events leading to the occurrence of a test pulse on the oscilloscope can be considered as having its genesis in one particular pulse (pulse A, waveform 2) in the train of pulses leaving the two-diode pulse-generator stage.

This particular pulse, occurring approximately once every 64  $\mu$ s has a consequent history which includes progression through the NAND gate, expansion in time and eventual selection as trigger pulse for initiating the test pulse.

Hence it is this pulse and the corresponding timing pulse derived from it (pulse B, waveform 3) which are alone among their respective pulse trains in having constant time displacements  $t_1$  and  $t_2$  from the leading edge of the test pulse (waveform 10) which are independent of frequency.

With the length of coaxial cable joining the timing wave to the oscilloscope chosen to have a propagation delay equal to the time displacement  $t_2$ , the design requirement of continuous coincidence of the timing pulse (pulse C, waveform 11) with the leading edge of the pulse being measured is thereby achieved.

### 3. Performance

The system as described and implemented is suitable for pulse-duration measurements in the range 20 ns to about 10  $\mu$ s. The latter value is somewhat arbitrary, being a practical rather than theoretical limit imposed by the difficulties of displaying and viewing the narrow timing pulses using long time scales.

The instrument can be used with any variable frequency oscillator having a sinusoidal output in the range 100 kHz to 50 MHz and of between 4 V and 6 V peak amplitude when terminated in 75 ohms.

The peak amplitude of the timing pulses at the oscilloscope depends on the length and attenuation of the delay line used for a particular measurement; with zero length of line, the peak amplitude, measured with a 5 V sinusoidal input waveform and timing-pulse output correctly terminated in 50 ohms, is  $-1.5$  V; the variation in timing-pulse amplitude with repetition period for any fixed length of delay line, is not greater than 2% throughout the range 20 ns to 10  $\mu$ s.

The peak amplitude of the 30 ns, 15.6 kHz trigger pulses, with the trigger output correctly terminated in 50 ohms, is 2 V and the maximum possible variation in trigger frequency from the nominal 15.6 kHz is  $-1000/64(1+2f)$  kHz for any input frequency,  $f$  MHz.

For measurements of rectangular pulse duration in the range 100 ns to 10  $\mu$ s the accuracy of the method is better than 0.1%. In the range 20 ns to 100 ns, pulse durations are resolvable to within  $\pm 100$  ps.

In the case of pulse-bandwidth measurements, overall accuracy is largely determined by the precision with which the 1:0.8727 amplitude ratio can be ascertained. Using an ancillary amplitude calibrator to provide suitable horizontal reference lines on the oscilloscope against which the test-pulse responses can be aligned, determinations of pulse bandwidth can be obtained with accuracies of about  $\pm 1\%$  or better.

The range over which measurements of pulse bandwidth are possible is basically 25 MHz to 125 kHz, the low-frequency limit being set for any particular system under test when the separation between consecutive test pulses (nominally 64  $\mu$ s) is insufficient to prevent noticeable overlap in the responses of the system to adjacent test pulses.

### 4. Conclusions

The equipment described offers a precise means of determining the duration of pulses longer than 20 ns emanating from sources which can be triggered at a nominal 15.6 kHz rate. In particular, in the field of waveform transmission systems, it fulfills the need for a convenient and accurate method of pulse-duration measurement required in pulse-bandwidth measurements, for which purpose the equipment is primarily intended.

The operation of aligning sharp timing pulses to the edges of the pulse being measured which forms the basis of the measurement, has been made as simple as possible. In this respect, permanent registration of a timing pulse with one pulse edge during alignment, negligible variation of timing-pulse amplitude, and stability of display without control, are the noteworthy features.

The technique has been implemented principally by the use of discrete component circuitry but certain integrated circuits available since the time of design and utilizing high-speed emitter coupled logic could perhaps now be employed. However, such modification, although in keeping with modern trends, will not obviate the use of discrete components entirely and is unlikely to offer any significant benefits in performance or enhanced accuracy.

### 5. Acknowledgment

Acknowledgment is made to the Senior Director of Development of the Post Office for permission to publish this paper.

### 6. References

1. Lewis, N. W., 'Television bandwidth and the Kell factor', *Electronic Technology*, 39, pp. 44-47, February 1962.
2. Lewis, N. W., 'Optimisation of lowpass systems for analogue signals', *Electronics Letters*, 3, pp. 199-201, May 1967.
3. St. Marie, L., 'Fast pulse generator is temperature stable', *Electronics*, 40, pp. 70-71, 6th February 1967.
4. Macdiarmid, I. F. and Phillips, B., 'A pulse and bar waveform generator', *Proc. Instn Elect. Engrs*, 105, Part B, pp. 440-448, 1958. (I.E.E. Paper No. 2687R September 1958.)
5. Patent No. 2,663,806, Darlington assigned to Bell Laboratories.

*Manuscript first received by the Institution on 4th December 1970 and in final form on 18th March 1971. (Paper No. 1405/IC49.)*

© The Institution of Electronic and Radio Engineers, 1971

# The Television Performance of the Klystron Amplifier

By

C. J. EDGCOMBE, Ph.D.†

and

C. N. O'LOUGHLIN, B.Sc.†

The paper describes the electrical behaviour of multi-cavity klystrons, now extensively used as high-power amplifiers in u.h.f. television transmitters. The factors affecting the frequency response are reviewed, with descriptions of the particular conditions required for output coupling and cavity tuning. The response of the klystron to the television tests for non-linearity is discussed in detail, and it is suggested that (a) differential phase shift is produced by the basic a.m.-p.m. characteristic of the tube; and (b) the discrepancy between observed results of line-rate linearity and differential gain tests is due to differing amplification of single and double sidebands, to be expected with any non-linear transfer characteristic. This discrepancy can be removed by suitable signal processing in the transmitter.

## 1. Introduction

Over seven hundred multi-cavity klystrons are now in service as high-power u.h.f. amplifiers for European television networks, and more than a million running hours are planned, during 1971, for the klystron complement of one user alone. The klystron (Figs. 1 and 2) is the usual choice for this type of service, because it has been shown to provide the necessary performance, in a reliable way, during the many thousands of hours of an individual tube's operation. The reliability and stability of performance, resulting from the tube's rugged construction, are usually achieved without difficulty when the tube has been properly set up on installation. As operating experience has grown, however, discussions have been held from time to time between klystron users and manufacturers, to determine the most appropriate tuning procedures for television service. These discussions have shown that particular facets of the tube's behaviour could usefully be described in more detail than has so far been published; also, the responses to the standard television test waveforms have been found to

differ from the ideal results in ways that would not be expected from the single-frequency gain characteristic. It appears that the non-linear behaviour of the klystron modifies the components of the measuring waveforms by means that have not been extensively discussed.

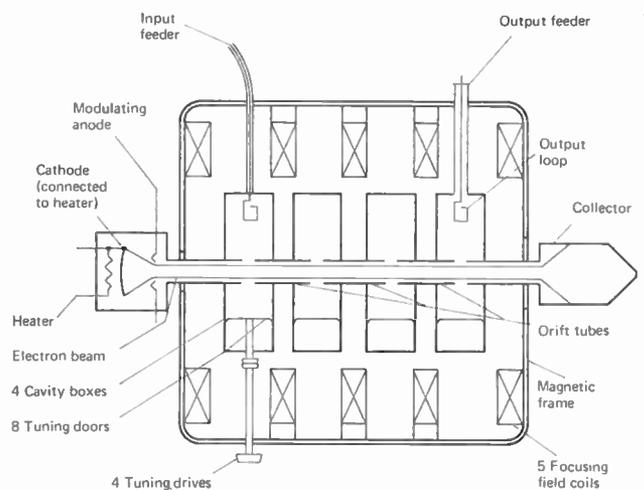


Fig. 2. Schematic diagram of 4-cavity klystron amplifier.

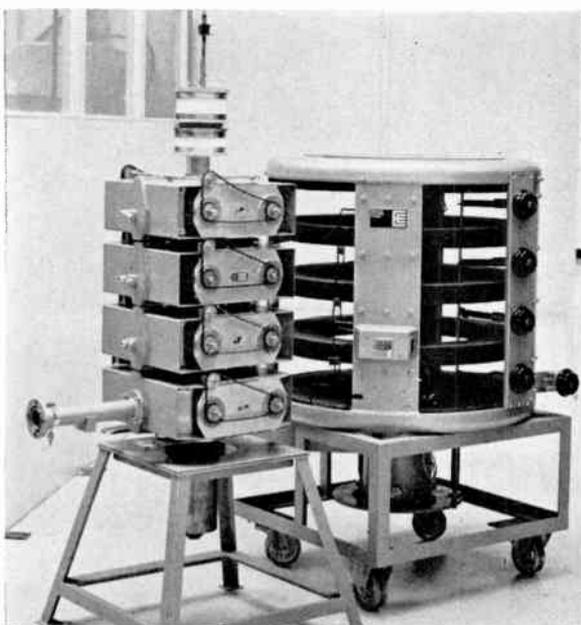


Fig. 1. Klystron and focus mount for 11 kW u.h.f. transmitters.

The purposes of this paper are thus: (1) to describe certain features of the behaviour of klystrons, in order to explain the tuning procedure normally adopted; and (2) to suggest how the non-linear transfer characteristic may modify the measuring waveforms so as to produce the results observed. The authors hope that the descriptions will be of use to both designers and operators of transmitters using klystrons; as the work given here applies chiefly to klystron behaviour, it does not deal with other possible sources of distortion of the television signal. It is suggested that the vestigial-sideband character of the signal is responsible, in conjunction with the klystron's non-linear behaviour, for producing the effects discussed, and also that the technique of modulation used can influence the choice of circuitry for correcting the signal distortion. Since the asymmetry of the signal

† English Electric Valve Co. Ltd., Chelmsford, Essex.

sidebands affects their amplification, it may be desirable in equipment design to consider the behaviour of components used for shaping the response, in conjunction with the characteristics of the klystron.

## 2. The Frequency Response of the Klystron

### 2.1 The Multi-cavity Gain Mechanism

The bunching effect that is produced when an electron beam is modulated by an alternating voltage has been fully described in many textbooks (e.g. refs. 1, 2). It can be summarized here by noting that velocity modulation, initiated at the drift-tube gap within the input cavity, causes the electrons progressively to group themselves into bunches, so that, at a suitable distance away from the first gap, the beam can induce current into a second cavity and generate a corresponding voltage across its gap. The cavity behaves as a parallel-resonant circuit, and, if driven at a frequency near resonance, the voltage set up can be many times the previous modulating voltage. The beam is then modulated again, more strongly than it was initially (Fig. 3), producing a still greater r.f. current and voltage at the third gap. The process can be continued as far as necessary by adding further cavities, spaced at suitable intervals along the

beam. In the final gap, the maximum r.f. current that can be produced is of the same order of magnitude as the d.c. beam current, so that power is extracted under large-signal conditions. The optimum drift length is then less than in the earlier stages, which are effectively small-signal amplifiers.

Although current modulation due to the first gap remains on the beam after the second and subsequent gaps, its effect in television service is likely to be much smaller than that of the remodulation. The response of the complete klystron to tuning adjustments then resembles that to be expected from a multi-stage i.f. amplifier, with single-tuned resonant circuits at input and output and between each amplifying stage, and with the output circuit heavily loaded to achieve maximum power transfer. The reason why such comparatively heavy loading is normally applied to the klystron output circuit, and its effect on the frequency response, are described in the following two sections.

### 2.2 Output Cavity Loading and Klystron Efficiency

It was stated in the previous Section that the r.f. current developed on the electron beam in the output gap may be of the same order as the d.c. beam current,  $I_0$ . The maximum available current depends on many

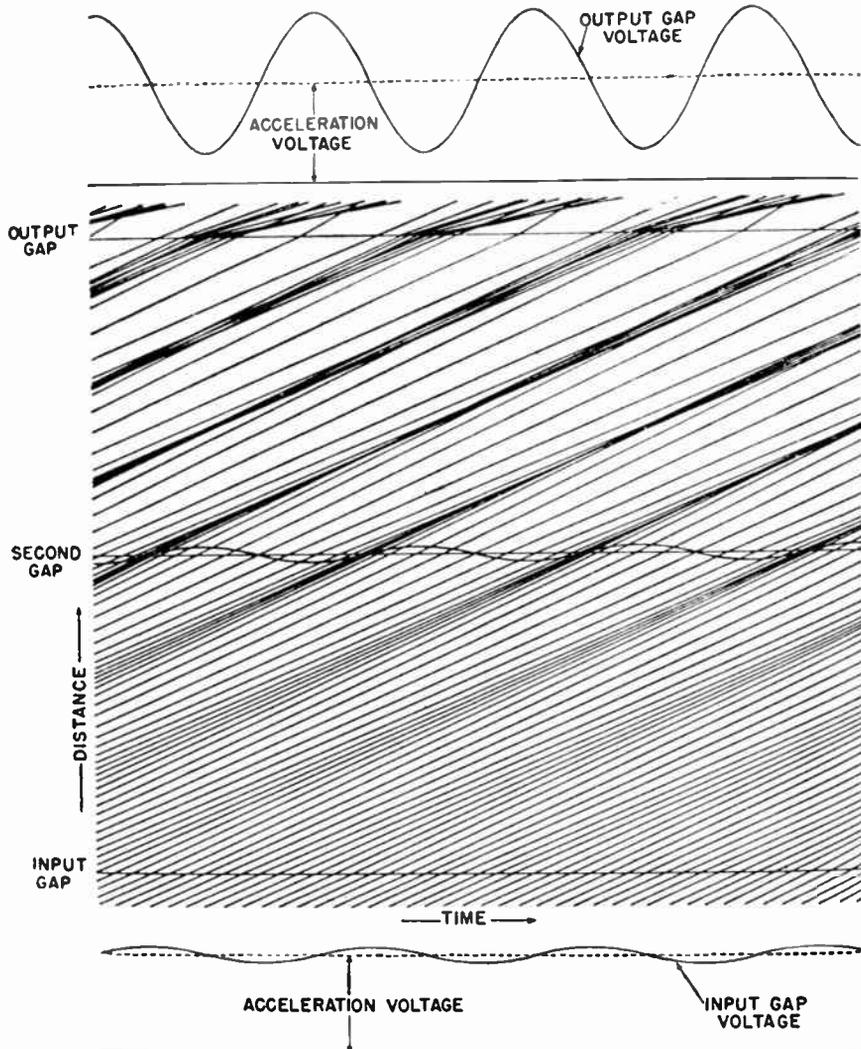


Fig. 3. Applegate diagram for 3-cavity klystron (After Harrison, A. E., 'Klystron Tubes', (McGraw-Hill, New York, 1947).

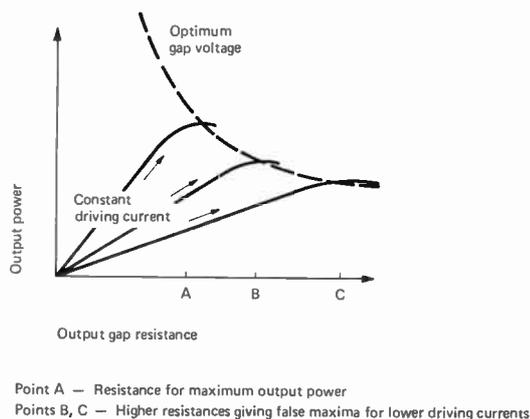


Fig. 4. Variation of output power with output gap resistance.

factors, but in practice its peak value is probably in the range  $0.8-1.2I_0$ . This indicates that there is an upper limit to the r.f. current that can be extracted under given conditions of tuning and loading, however much drive power is available.

There is also an optimum value for the r.f. voltage developed at the output gap. If the r.f. driving current is kept constant, and the output gap resistance increased from its minimum value (by reducing the output coupling), the power transferred to the output circuit initially increases (Fig. 4). At some value of coupling, however, the gap voltage becomes large enough to stop just the slowest electrons; if the gap resistance is increased still further, the gap voltage tends to increase, so that more electrons are stopped, and then accelerated backwards out of the gap, absorbing useful r.f. power. Thus the power coupled out of the output cavity reaches a maximum when the peak gap voltage has a particular value, generally considered to be close to the beam voltage,  $V_0$ .

The separate limitations on both the r.f. current and r.f. voltage that can usefully be developed in the output cavity indicate that: (1) the maximum output power is achieved when both limitations are reached together, giving a narrow-band efficiency in the range of 40-50% for the types of tube now in service; and (2) the maximum efficiency is achieved when the load impedance, as seen at the output gap, is adjusted to the ratio of the maximum useful r.f. voltage to the maximum achievable r.f. current. This impedance is approximately equal to the beam impedance,  $V_0/I_0$ , for both narrow-band and broad-band operation, although of course in television service the numerical values of beam impedance are appreciably different in sound and vision amplifiers. The optimum output impedance may in practice also be affected by the output circuitry, so that tubes intended to work over many television channels and power levels can be run most efficiently when the output coupling is adjustable.

In order to set up the output loading in practice, it may be found convenient first to set the coupling to its maximum value. The drive power is then increased until the output power appears to saturate—this indicates that the r.f. current has reached its maximum value. The output coupling is now reduced steadily, without changing the drive level, and the power output is found

to increase correspondingly to a second maximum. When this second maximum is reached, both r.f. current and r.f. voltage are optimized, and the coupling is correctly set for maximum efficiency. Unless some procedure of this sort is followed, it is possible to achieve the optimum gap voltage without developing the full r.f. current; then, increasing the drive level merely reduces the output power, because the increase in gap voltage turns back many electrons and sharply reduces the effective r.f. current. In this condition, the power output will not be the maximum of which the tube is capable.

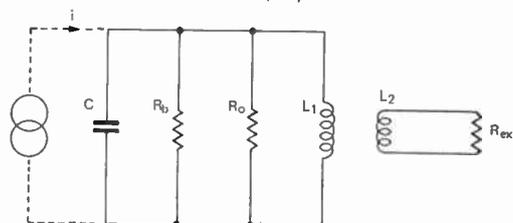
While discussing the subject of efficiency, it may be mentioned that the penultimate cavity is always tuned to the high-frequency end of the required pass-band, in order to present an inductive impedance at any driving frequency. For reasons which depend on the detailed mechanism of bunching, the efficiency so obtained is appreciably greater than if the penultimate impedance is made capacitive, i.e. than if the penultimate cavity is tuned below the driving frequency.

### 2.3 Cavity Resonators—Shunt Resistance, $Q$ -factor, and $R/Q$

Each of the cavity resonators behaves effectively as a parallel-tuned resonant circuit, with several possible sources of resistive loading (Fig. 5). In the earlier cavities, loading is provided principally by the beam, with a further contribution, if required, from inductively coupled external loads; tubes of current design, with a microperveance of 2, need much less external loading than earlier designs which had a microperveance of 1. The output cavity is of course principally loaded by inductive coupling to the output feeder and aerial. In all cavities, the finite conductivity of the silver-plated resonator surface produces a small further loss, which in u.h.f. tubes can usually be ignored. The shunt resistances which correspond separately to beam loading, external loading, and the cavity loss can be regarded as acting in parallel across the drift-tube gap. A cavity and its adjustable coupling loop behave as windings of a transformer with variable coupling, so that with a fixed external load a wide range of impedance values can be presented to the gap by varying the coupling.

In a low-frequency resonant circuit containing a parallel combination of inductance  $L$ , capacitance  $C$ , and shunt resistance  $R$ , the circuit  $Q$  can be defined by

$$R/Q = \sqrt{L/C}$$



- $i$  = R.F. current induced by beam
- $C$  = Capacitance between drift tube noses
- $R_b$  = Beam loading resistance
- $R_o$  = Cavity shunt resistance
- $L_1$  = Inductance of cavity box
- $L_2$  = Inductance of coupling loop
- $R_{ext}$  = Aerial or other external loading

Fig. 5. Equivalent circuit of klystron cavity.

so that the ratio  $R/Q$  is independent of the shunt resistance  $R$ . Thus for any change in  $R$ , there is a proportionate change in the circuit  $Q$ . The same property holds for resonant cavities, although the values of  $L$  and  $C$  can no longer be unambiguously identified; the value of  $R/Q$  is effectively a constant determined by the geometry of the cavity. It can be regarded as the 'characteristic impedance' of the cavity, and in klystrons its value is typically of the order of 100 ohms.

It follows from this property of the cavity that, when the coupling is adjusted to provide the correct resistance at the output gap for maximum efficiency, the  $Q$ -factor of the output cavity is also determined. For a klystron of the type now in service, with perveance  $2 \times 10^{-6} \text{ A/V}^{3/2}$ , the beam impedance when running at 12.5 kV, 2.85 A is 4400  $\Omega$ , and with an  $R/Q$  value of 100  $\Omega$ , the loaded  $Q$  at maximum efficiency will be of the order of 44. The impedance of the output cavity then varies by less than 40% over the bandwidth of one television channel; its response can be observed, if desired, by using a swept-frequency c.w. source to drive the tube to (current) saturation, so that the driving current is constant over the passband.

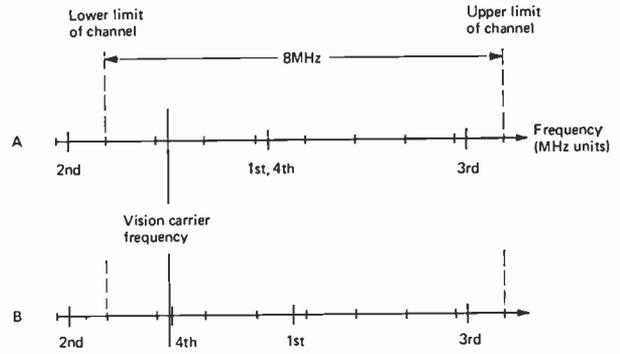
During the period of picture information in the television signal, the output level remains well below the saturated output of the klystron and the overall frequency response must be flat within specified limits. This is achieved by tuning and loading the earlier cavities so that the frequency variation of the current they deliver complements the output cavity response, over the passband. The values of  $R$  and  $Q$  used in the earlier cavities are in practice higher than those for the output cavity, because the r.f. currents are lower and the gap voltages are unlikely to approach the beam voltage; the earlier cavities can thus be adjusted to optimize the voltage gain, which in the earlier stages is more important than power gain.

The impedance of a parallel-resonant circuit, with the same component values as before, and with resonant frequency  $\omega_0$ , can be written as

$$Z = \frac{R/Q}{\frac{1}{Q} + j\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right)}$$

This shows that at frequencies several 3-dB bandwidths away from resonance, where  $1/Q$  is small compared with the frequency deviation, the impedance depends principally on the  $R/Q$  of the circuit, and very little on its  $Q$ . In practice this means, as is well known, that when the loading of a cavity is altered (changing  $Q$  but not  $R/Q$ ), the overall response varies most near the resonant frequency of that cavity.

For broadband operation, the voltage gain-bandwidth product of each stage is proportional to the cavity  $R/Q$ , which is a constant fixed in manufacture. The power gain of the complete multi-stage tube thus varies very rapidly with the reciprocal of the bandwidth. This explains the usual recommendation to keep the penultimate cavity tuned higher than the driving frequency, as, if it is inadvertently brought into line, the gain may rise



Note: Optimum tuning position for 3rd cavity depends on video bandwidth in use.

Fig. 6. Schematic tuning patterns for 4-cavity vision amplifier klystrons.

to such a high value that feedback round the associated circuitry can lead to oscillation.

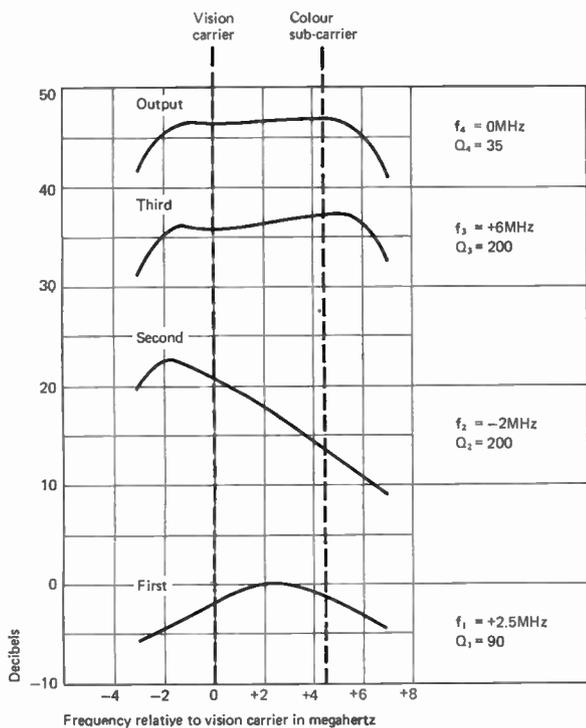
### 2.4 The Tuning Pattern

It has previously been suggested that the klystron can be considered for tuning purposes as a simple stagger-tuned amplifier, with special conditions for the penultimate cavity tuning and output loading. If these conditions are met, the tuning and loading of other cavities can be varied within wide limits to achieve the required overall response, with little effect on efficiency, though perhaps some change in gain.

Two tuning patterns that are frequently used are shown in Fig. 6. In the first of these, the cavity resonant frequencies are symmetrically distributed about the centre of the vision signal passband, and it is usually easy to obtain a sufficiently flat response, using a c.w. swept source. Where the best possible efficiency is required, however, the tuning pattern used is often more like the second one shown. This pattern is chosen because most of the power in the signal is radiated at vision carrier frequency, which is a few megahertz below the channel centre. The maximum output power is obtained when the output cavity is correspondingly tuned below the channel centre. It is then necessary to tune the input cavity to a frequency slightly above the channel centre, in order to keep the overall response suitably flat.

In order to see how the power levels at successive cavities vary with frequency, it is useful to have an approximate method of determining the response of the whole tube. The basic theory of klystrons can be used to show that, for a signal of fixed amplitude, at a single input frequency, the r.f. current produced by bunching is almost independent of frequency, over the bandwidth of one television channel.

The frequency response of the tube, and also its group delay variation, are therefore controlled almost entirely by the characteristics of the cavities. At linear levels of operation a working approximation to the overall frequency response can be obtained by multiplying together, at each frequency of interest, the responses of the cavities at that frequency, and this is easily done by addition when the scales are plotted in decibels. A typical result, showing how the relative power levels vary with frequency, for the second type of tuning pattern described above, is shown in Fig. 7.



Vision carrier frequency ~600 MHz, small-signal conditions.

Fig. 7. Typical frequency variation of signal levels at successive gaps in a 4-cavity klystron.

In discussions about tuning, the question is sometimes raised whether the input cavity should appear matched to the drive impedance. It has been shown above that the input cavity is usually tuned a few megahertz above the vision carrier frequency, round which most of the signal power is concentrated, and the working  $Q$  is determined by the need for a flat frequency response. Now it can certainly be arranged in designing the klystron that, over at least part of the tuning range, the  $Q$  required for matching is near that required for a flat frequency response. However, the cavity can only absorb an incident signal completely if its resonant frequency is identical with the frequency of the signal and, since the cavity is normally tuned to a frequency above the vision carrier, some of the carrier power is bound to be reflected. In normal operation, therefore, one should not expect the input cavity to present a perfect match to the drive.

### 3. Non-linear Amplification in Television Systems

#### 3.1 Operation of the Klystron with a Single-frequency Input

The general features of the non-linear behaviour of the klystron, when driven with a single-frequency c.w. signal, are well known. As the input power is increased from zero towards the saturation value, the ratio of output power to input power decreases (Fig. 8), becoming several decibels less at saturation than the small-signal value. There is also a variation of output phase with signal level, known as 'a.m. to p.m. conversion'; the phase shift is calculated<sup>3</sup> to be approximately proportional to input power and to the square root of beam

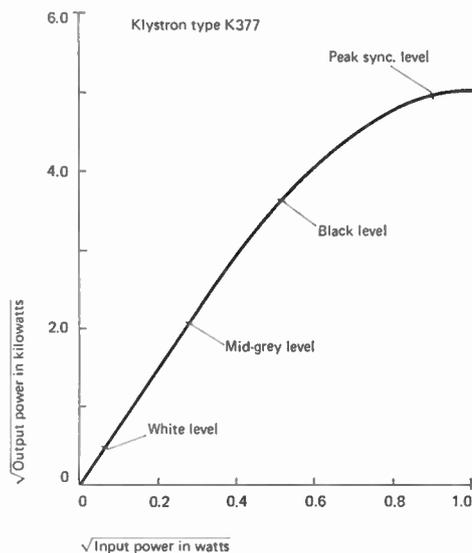


Fig. 8. Typical klystron transfer characteristic.

perveance, and to be about  $25^\circ$  at saturation for a klystron of microperveance 2.

It has previously been suggested<sup>4</sup> that a four-cavity klystron can be regarded as a two-stage linear amplifier, driving a final stage in which most of the non-linear behaviour occurs. This suggestion has been confirmed by the test results shown in Fig. 9. The figures are traces of waveforms obtained at successive cavities, using the standard television tests of line-time linearity and differential gain (to be discussed in detail later); although these are not strictly single-frequency signals, they do indicate that most of the non-linear behaviour occurs in the final stage, and that the previous stages are much more nearly linear in comparison. It can also be seen that the sub-carrier amplitude is reduced below that of the vision carrier at the second cavity and slightly increased at the third, as suggested by the calculated responses shown in Fig. 7.

The variation in single-frequency gain with signal amplitude need not be particularly frequency-dependent; the only frequency-sensitive components of the tube, the resonant cavities, are linear devices, while the bunching process in the electron beam is nearly independent of frequency, as stated in Sect. 2.4. It is quite

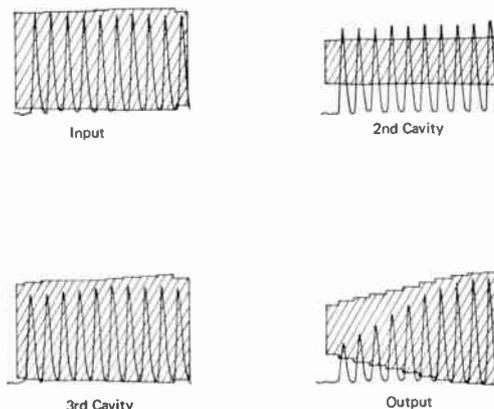


Fig. 9. L.f. linearity and differential gain responses.

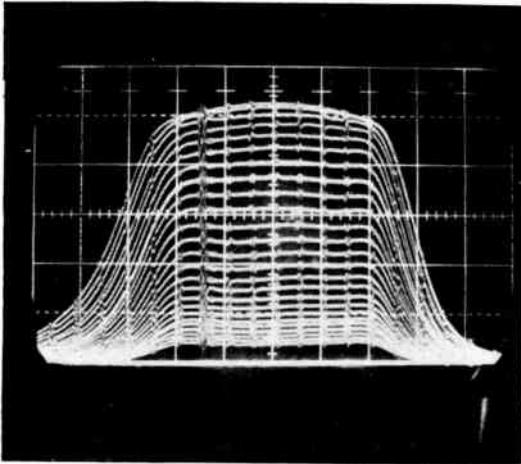


Fig. 10. Response of klystron to swept-frequency c.w. signal.

possible to tune a klystron so that its frequency response to a swept c.w. signal is flat, e.g. to within 1dB over an 8 MHz passband, at all output levels from saturation down to 20 dB or more lower. Tests of this sort are carried out on every klystron produced for television service (Fig. 10). This point is emphasized here, because initial experience with transmitters for colour television appeared to suggest that the non-linear behaviour does vary with frequency. It is an aim of this paper to show that the apparent behaviour is connected with the vestigial-sideband character of the television signal, and that suitable signal processing can counteract the observed frequency dependence.

### 3.2 Linearity Tests in Television Systems

In assessing the performance of television systems, it is necessary for all items of equipment in the transmission path to be subjected to the same standardized tests, as specified by the broadcasting authorities. It has therefore been desirable to understand klystron performance, as well as to discuss it, in terms of these particular test signals. The particular tests concerned are those of line-time (or low frequency) linearity, differential gain and phase and sideband response; details of the types of waveform specified for international use can be found in references 5 and 6.

At this stage it is worth pointing out one difference between the line-time linearity and differential gain tests.

The video frequency signal used for differential gain measurements consists of a low-frequency component, changing in amplitude during the period of one line, together with a sub-carrier at a relatively high frequency (typically 4.43 MHz), on which measurements are made after demodulation. On the other hand, the information obtained from the line-rate test is contained almost completely in a smaller frequency range, within a few hundred kilohertz of the carrier frequency. The difference between the frequencies used in the two tests is significant, because the vestigial-sideband nature of the r.f. signal (Fig. 11) ensures that only a single sideband of any video frequency above about 1½ MHz is transmitted, while for all low-frequency components of the signal, both sidebands are retained.

### 3.3 The Response of the Klystron to Television Test Signals

When the test signals used for colour television are applied, via a modulator, to a klystron amplifier, two particular characteristics of the behaviour appear to need further explanation. One of these is the differential phase response; it is not immediately obvious why the phase of the sub-carrier should alter, as the carrier amplitude changes during one line. The second effect is that while both the line-time linearity and differential gain tests show a general reduction in gain with increasing carrier level, as expected, the results found in the two tests are often different, with the differential gain waveform usually showing less compression. This discrepancy has to be considered when the design of correction circuitry is being undertaken for transmitters.

Theoretical work on the first of these questions, outlined in Appendix 2, suggests that the differential phase shift arises because of the existence of a.m.-p.m. conversion, as mentioned in Section 3.1. The value of differential phase shift at a particular carrier level is calculated to equal the phase shift produced in the carrier itself, at the same level, by a.m.-p.m. conversion. The differential phase shift can thus be expected to vary, during the period of the test waveform, approximately as the square of the carrier amplitude. If the ratio of sync power to saturated power is taken as 0.9, and the black/sync power ratio as 0.58, the phase shift between black level and white level is calculated to be about 8°, for a klystron of micro-perveance 2. This work has not indicated any direct connexion between the curvature of the transfer charac-

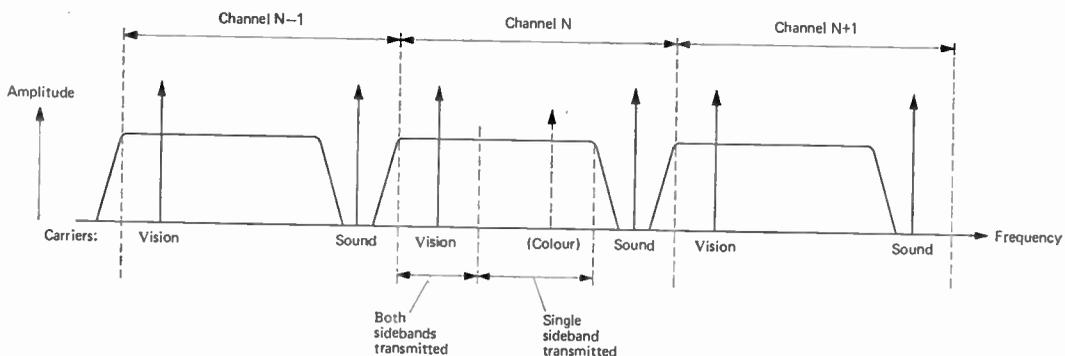


Fig. 11. Typical vestigial-sideband response of television transmissions.

teristic and the phase shifts, nor does it suggest that a.m.-p.m. conversion has any first-order effect on line-time linearity or differential gain results.

The second problem, of the discrepancy between line-time linearity and differential gain results, has been clarified by practical work, using variations on the standard tests. Measurements of differential gain on a single transmitter, using other video frequencies than that of the colour sub-carrier, indicated that the change in non-linear behaviour took place at a video frequency between 1 and 2 MHz. The same frequency is characteristic of a notch or 'step' that is sometimes observed in measurements of the sideband response of the klystron. When this sideband step was examined, the frequency at which it occurred was found to change as the low-level vestigial-sideband (v.s.b.) filter (normally fitted between modulator and klystron) was connected into or out of circuit. Further tests, in which the low-frequency limit of the passband was altered by means of the cavity tuning, with the v.s.b. filter out of circuit, showed that the step in response always occurred at the same frequency interval above vision carrier as the separation between vision carrier and the lower limit of the passband. It therefore appeared that some property of the signal was involved, since the step frequency could be made to change by removing the v.s.b. filter, without touching the klystron.

This behaviour, together with the vestigial-sideband nature of the transmitted signal (Fig. 11), suggests that the sideband compression is in fact greater in the frequency range where both sidebands are retained, than it is where only one sideband is amplified. In the next section a possible explanation is proposed for this effect.

### 3.4 Single and Double Sideband Amplification in Non-linear Devices

The behaviour will be described here qualitatively, using vectors to represent the test signal, as this method seems simplest to appreciate in general terms. The results given have been justified mathematically, both in detail using the ballistic theory of the klystron, and also more generally, by means of a simple model of non-linear behaviour containing the essential points mentioned in Section 3.1. Details of these calculations are given in Appendix 1.

Consider a carrier that has been amplitude-modulated to produce two sidebands (Fig. 12(a)), and suppose first that both sidebands are retained in subsequent non-linear amplification. The combination of the two sidebands and the (fixed-amplitude) carrier can equally well be regarded as a signal vector fluctuating in amplitude, and only in amplitude, at the video frequency used for modulation. Provided that the transfer characteristic measured for c.w. signals still represents the behaviour of signals varying at video frequency, the effect of saturation is to reduce the ratio between sideband and carrier amplitudes at the output, compared with their ratio at the input to the device. The sidebands can be completely suppressed if their initial amplitude is small and the carrier amplitude is that required for saturation. This suppression has been observed by using the line-time linearity test (Fig. 13).

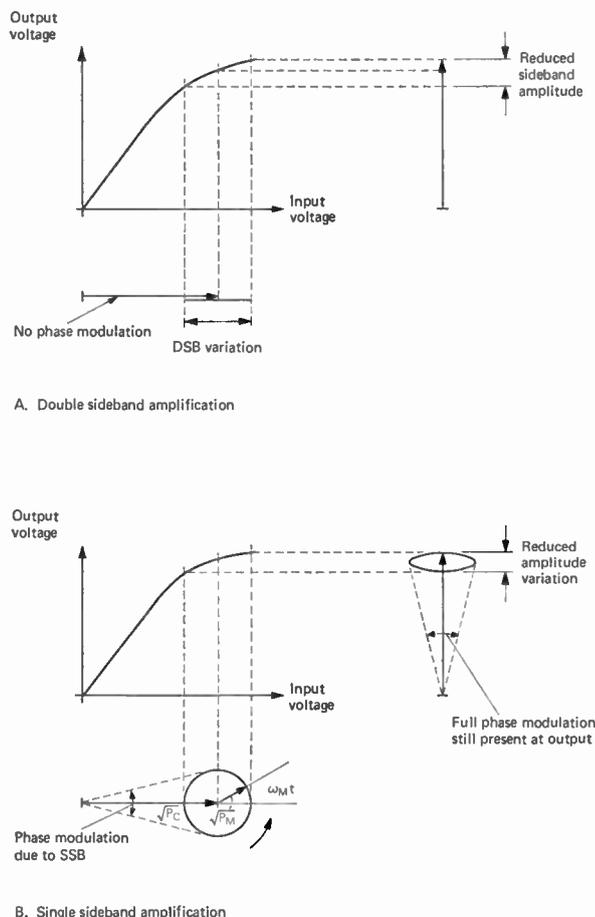
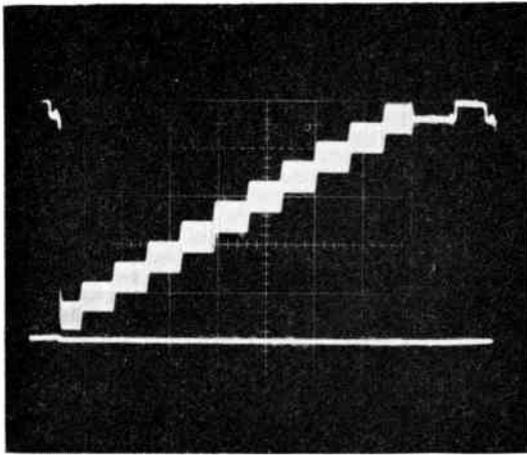


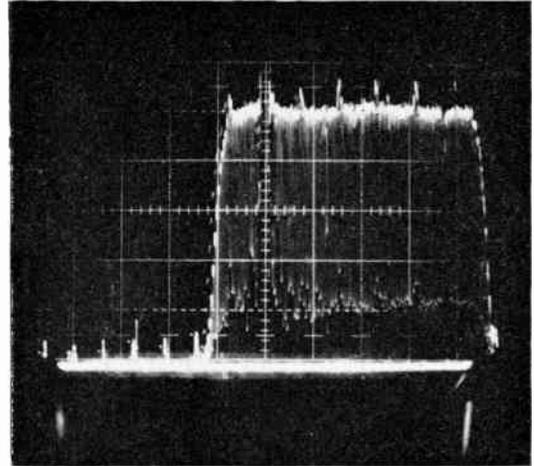
Fig. 12. Non-linear amplification of carrier and double or single sidebands.

Now suppose that the same signal has its lower sideband removed by filtering, before being passed through the non-linear amplifier. The remaining signal can be regarded as fluctuating in amplitude, as before, but also now in phase, relative to the unmodulated carrier. When this signal is passed through the amplifier characteristic of Fig. 12(b), the amplitude modulation is reduced or even eliminated, but the phase fluctuation is unaffected. Then even at saturation, the amplified signal retains a phase variation at video frequency, and so can be resolved as usual into a carrier and two contra-rotating sidebands. When the carrier amplitude is large enough to saturate the amplifier, therefore, a single additional sideband is not completely crushed, and the complementary sideband is regenerated; whereas if both sidebands are initially present as amplitude modulation, they are completely suppressed at saturation. Although in television service the carrier amplitude is usually lower than the saturation value, the compression of single and double sidebands is still different, because the locus of the tip of the single-sideband vector is compressed to an ellipse, instead of the circle corresponding to each of the double sidebands.

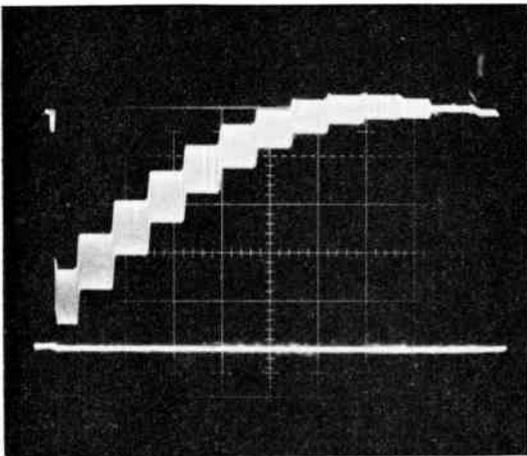
The reduced compression of a single sideband corresponds to the behaviour seen in the differential gain test. The modulator produces both sidebands from the video input signal, but the lower sideband is normally removed,



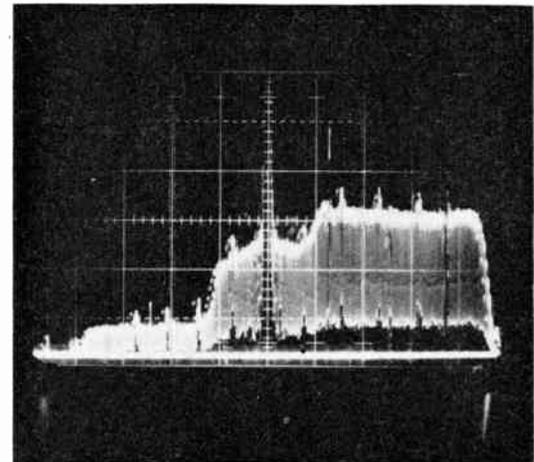
(a) Input waveform



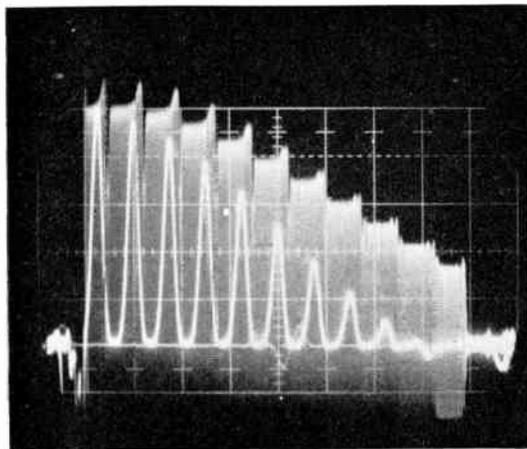
(a) Picture level 15% of sync level.



(b) Output waveform

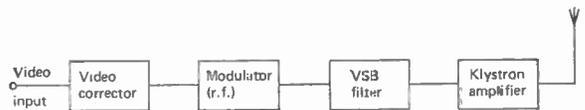


(b) Picture level 80% of sync level.

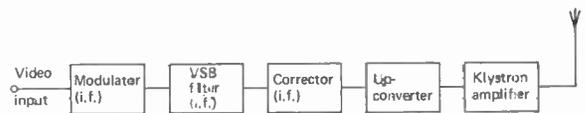


(c) L.f. linearity and differential gain responses.

Fig. 14. Sideband response of klystron at two carrier levels. Vertical axes, voltage; horizontal axis, frequency.



A. Modulation direct to r.f.



B. Modulation to i.f.

Fig. 15. Skeleton arrangements of transmitter circuitry.

Fig. 13. Response of klystron driven to saturation. Vertical axes voltage; horizontal axis, time (increasing from right to left).

before reaching the klystron, by the v.s.b. filter. Even if the filter is not in circuit, the lower sideband will still be removed because it lies outside the passband defined by the earlier klystron cavities. The regeneration of the lower sideband takes place subsequently, however, and so may be noticeable in the output signal (Fig. 14), because the response of the final cavity is relatively broad-band.

Figure 13 shows the behaviour of a klystron when driven to saturation with the differential gain waveform. The output signal has been passed through a demodulator with the usual vestigial-sideband response, and so retains only the wanted sideband. The steps in carrier amplitude, showing the linearity response, have been reduced completely to zero; in contrast, the sub-carrier sideband is only partly crushed at saturation. The sideband amplitude in this particular case has dropped at saturation to about a third of its original size; the theory given in Appendix 1 suggests a factor of a third, but is only to be taken as a rough indication because the model used cannot be expected to be accurate right up to saturation.

### 3.5 Non-linear Behaviour in Television Service

The previous section has suggested that single and double sidebands are compressed to different extents by a non-linear amplifier, in the presence of the carrier. If the klystron amplifier is operated without any pre-correction for non-linearity, the difference in sideband compression can be observed by displaying the sideband analyser response and tuning the klystron for a flat response near white level; then, as the picture level is raised towards black, the step will be seen to appear at the frequency corresponding to the change from single to double sidebands. (Fig. 14).

From what was said in Section 3.1 about the linear behaviour of the earlier stages of the klystron, it is clear that altering the cavity tuning is likely to change the gain of these stages far more than their linearity. The linearity of operation of the final stage will only be affected if the signal amplitude alters in the non-linear region. Thus in transmitter service, where it is usually necessary to work at a fixed output level, variation of the tuning pattern should not be expected to have very much effect on the linearity of klystron operation.

While this paper shows how a non-linear transfer characteristic can produce a sideband step, it has also been suggested<sup>7</sup> that a similar effect may be produced by the phase characteristic of the v.s.b. filter. Some care may thus be needed, in practice, in deciding how far each component of a transmitter is responsible for any distortion that is observed overall. The arguments presented here are not limited to microwave tubes, and so similar behaviour may be expected in any part of the system where non-linear behaviour is possible, including for instance an incorrectly-operated sideband analyser mixer.

### 3.6 Linearity Correction in the Transmitter

In practice, television transmitters are provided with correction circuits to enable the overall response of the amplifier to be made linear in amplitude and phase.

Two arrangements of modulator and correction circuitry are currently in use (Fig. 15):

- (a) Correction at video frequency, followed by modulation directly on a vision carrier at the required final frequency, and then by frequency response shaping by means of the v.s.b. filter.
- (b) Modulation on some convenient intermediate frequency carrier, followed by response shaping, then by correction, and by conversion to the required final frequency.

With the first of these methods, the modulator produces a signal containing both sidebands of all video frequencies, and the v.s.b. filter removes the lower sideband over a limited range. If correction were applied uniformly at all video frequencies, then the resulting signal would have the correction appropriate to double sidebands even in the single-sideband region, and the 'step' would still be present. In practice this means that the correction circuitry must operate differently in the two ranges of video frequency, the changeover frequency depending on the television standard in use. Any remaining irregularity in the amplifier response near the upper sideband of this frequency can often be reduced by attention to the klystron cavity loading, where this is adjustable.

In the second modulator arrangement, the frequency response is shaped before the signal reaches the correction circuit. The same number of sidebands then pass through the corrector as through the klystron for all video frequencies, except for a limitation of the regenerated sideband by the earlier cavities of the klystron. In practice, however, this limitation does not appear to have serious effects, and it is found sufficient for the corrector circuit to be relatively insensitive to frequency. The double sidebands are automatically stretched more than the single ones, by the inverse of the crushing effect described for the klystron.

## 4. Conclusion

For purposes of achieving the required frequency response, at output levels below saturation, the multicavity klystron can be regarded as a multi-stage amplifier with nearly independent inter-stage tuning, and with special conditions on the output coupling and penultimate cavity tuning. The differential phase shift observed in television service is calculated to occur as a result of a.m.-p.m. conversion in the klystron, and to be of similar magnitude to the carrier phase shift. The observed differences between the line-time linearity and differential gain characteristics of the klystron alone, and the step seen in its sideband response, are due to the differing compression of single and double sidebands, to be expected from any device with a non-linear transfer characteristic. These effects can be corrected by using frequency-insensitive circuitry following i.f. modulation and shaping, or by circuitry operating on the video signal, before modulation direct to the final frequency.

## 5. Acknowledgments

The suggestions put forward in this paper are the results of discussions among many engineers who have

been concerned with the problems of television transmission. We should particularly like to thank Mr. J. Dain and Mr. M. Esterson for their helpful comments. We are also grateful to the GEC-Marconi Co. and to Pye TVT Ltd. for the provision of experimental facilities. Finally, we wish to thank the Managing Director of GEC Electronic Tube Co. for permission to publish this paper.

**6. References**

1. Spangenberg, K., 'Vacuum Tubes' (McGraw-Hill, New York, 1948).
2. Beck, A. H. W., 'Thermionic Valves' (Cambridge University Press, 1953).
3. Edgcombe, C. J. and Heppinstall, R., 'Non-linear behaviour in klystrons for u.h.f. television.' Proc. 7th Conference on Microwave and Optical Generation and Amplification (Hamburg, 1968), p.151.
4. O'Loughlin, C. N., 'Intermodulation characteristics of high-power klystrons used in frequency transposers for colour tv,' Proc. Int. Conf. on U.H.F. Television, London, 1965, paper 10 (I.E.R.E. Conf. Proceedings No. 6).
5. Report of XIth Plenary Assy of C.C.I.R. (Oslo, 1966), Vol. V, Recommendation 451, pp. 92, 93, 98.
6. Ludbrook, R., 'Television transmitter sideband analyser', *Marconi Instrumentation*, 9, No. 1, p.13, 1963.
7. Hatfield, G. E., 'Improving performance from tv broadcast transmitters', Proc. Int. Broadcasting Convention, 1968, paper 9/6. (I.E.E. Conference Publication No. 46, Pt. I.)
8. Wallander, S. 'Non-linear TWT Analysis with Two Signal Frequencies', Research Lab of Electronics Report No. 59, Chalmers University, Gothenburg (1966).
9. Eaton, J. L. and Whythe, D. J., 'Comparison between measured and theoretical non-linear behaviour of 4-cavity klystrons', *Electronics Letters*, 3, No. 2, p.55, February 1967.
10. Wallander, S. 'Non-linear multi-signal twt theory'. *Internat. J. Electronics*, 29, No. 3, p.201, 1970.

**7. Appendix 1.**

**Mathematical theory of non-linear single and double-sideband amplification**

In order to discuss the non-linear behaviour of a klystron or other device, it is necessary to define the variation of both gain and phase with signal amplitude. Consideration of the ballistic theory of the klystron for a single frequency input signal<sup>3</sup> suggests that the voltage gain decreases from its small-signal value, and the phase lag through the tube increases, in both cases as the square of the signal amplitude. Such measurements as have been made suggest that this model is a reasonable approximation to the real behaviour of tubes, at least at small signal levels. We thus assume that, when a single-frequency signal

$$v_{in} = \sqrt{P} \cos(\omega t + \phi) \quad \dots\dots(1)$$

is applied to the input of a tube, the output signal can be represented by

$$v_{out} = G\sqrt{P(1-cP)} \cos(\omega t + \phi + kP), \quad \dots\dots(2)$$

where  $G$ ,  $c$  and  $k$  are constants for the device under consideration. The constant  $k$  is taken as zero in this Appendix, but can alternatively be used to deduce the effects of a.m.-p.m. conversion (see Appendix 2).

A model of this form has been suggested by Wallander<sup>8</sup> to summarize his detailed work on the non-linear behaviour of travelling-wave tubes; he has extended it

recently<sup>10</sup> to describe their amplitude non-linearity in more detail. The use of the same model to describe both klystrons and t.w.t.s suggests that it may be a useful general means of describing non-linear amplification in most microwave tubes, and possibly in other devices as well.

The particular value of this model is that  $P$  and  $\phi$  are presumably not limited to constants, provided that any fluctuations with time are slow enough. Since these parameters vary only at video frequency, results deduced by this means should be less unreliable than others obtained by forcing instantaneous value of the r.f. signal through a power series derived from the static characteristic.

The resolution of an amplitude-modulated signal into a carrier and sidebands is well known. If a carrier, of amplitude  $\sqrt{P_c}$  and frequency  $\omega$ , is amplitude-modulated with a tone of amplitude  $2\sqrt{P_m}$  and frequency  $\omega_m$ , the resulting signal can be represented by

$$\begin{aligned} v_2 &= (\sqrt{P_c} + 2\sqrt{P_m} \cos \omega_m t) \cos \omega t \\ &= \sqrt{P_2(t)} \cos \omega t \\ &= \sqrt{P_c} \cos \omega t + \sqrt{P_m} \cos(\omega - \omega_m)t + \\ &\quad + \sqrt{P_m} \cos(\omega + \omega_m)t \end{aligned} \quad \dots\dots(3)$$

The signal can thus be regarded as having a time-varying amplitude  $\sqrt{P_2(t)}$ , fluctuating at the (video) frequency  $\omega_m$ .

If the lower sideband is removed, the signal becomes instead

$$v_1 = \sqrt{P_c} \cos \omega t + \sqrt{P_m} \cos(\omega + \omega_m)t \quad \dots\dots(4)$$

which can be expressed as

$$v_1 = \sqrt{P_1(t)} \cos(\omega t + \phi(t))$$

where

$$\begin{aligned} P_1(t) &= P_c + P_m + 2\sqrt{P_c P_m} \cos \omega_m t \\ \phi(t) &= \sin^{-1} \left\{ \frac{\sqrt{P_m} \sin \omega_m t}{\sqrt{P_1(t)}} \right\} \end{aligned} \quad \dots\dots(5)$$

Thus the single-sideband signal can be regarded as a tone of frequency  $\omega$ , modulated in both amplitude and phase according to (5).

On replacing  $\sqrt{P}$  in (2) by the signal amplitude,  $\sqrt{P_2}$ , defined in (3), the amplitude of each sideband when the other is present is found (with  $k$  taken as zero) to be

$$G\sqrt{P_m} [1 - c(3P_c + 3P_m)] \quad \dots\dots(6)$$

However, if instead the single-sideband values of  $P_1$  and  $\phi$ , defined by (5), are inserted into (2), we find that the complete signal becomes

$$\begin{aligned} v_{1 \text{ out}} &= G(1-cP)(\sqrt{P} \cos \phi \cos \omega t - \sqrt{P} \sin \phi \sin \omega t) \\ &= G\{1 - c(P_c + P_m + 2\sqrt{P_c P_m} \cos \omega_m t)\} \times \\ &\quad \times (\sqrt{P_c} \cos \omega t + \sqrt{P_m} \cos(\omega + \omega_m)t) \end{aligned}$$

From this result it can be deduced that the amplitude of the output component at frequency  $(\omega + \omega_m)$  is

$$G\sqrt{P_m} [1 - c(2P_c + P_m)] \quad \dots\dots(7)$$

Comparison of (6) and (7) then shows that when the power in the sidebands is small compared with that of the carrier, i.e. when  $P_m \ll P_c$ , the reduction in amplitude

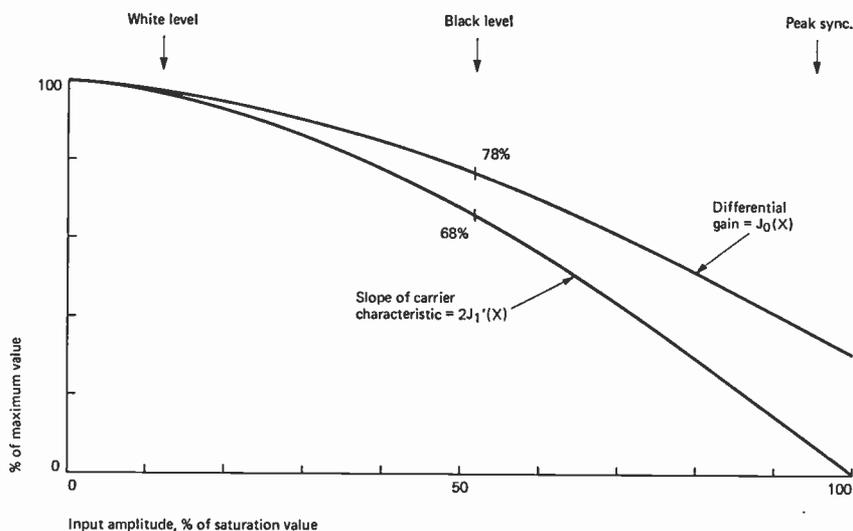


Fig. 16. Functions related to l.f. linearity and differential gain.

of a single sideband is only  $2/3$  of that of one of a pair of a.m. sidebands, whatever the value of  $c$ .

For comparison, it is worth quoting some results of the detailed theory of klystrons. It is well known that the two-cavity ballistic theory for a single driving frequency indicates an output current proportional to  $J_1(X)$ , where  $X$  is a parameter proportional to input voltage. The ratio of the slope of the gain characteristic at drive level  $X$ , to that at small drive levels, is thus  $J_1'(X)/J_1'(0) = (J_0(X) - J_2(X))$ . The l.f. linearity test provides an approximate measure of this ratio, using both sidebands of the carrier to transmit the information.

When two tones are present, however, at levels  $X_1$  and  $X_2$ , and frequencies  $f_1$  and  $f_2$  respectively, the output amplitude at  $f_2$  is proportional to  $J_0(X_1) J_1(X_2)$ .<sup>9</sup> Thus when  $X_2$  is held constant and  $X_1$  is changed, the output at  $f_2$  changes only as  $J_0(X_1)$ . This signal composition is characteristic of the differential gain test, where the sub-carrier amplitude is examined as the carrier level is altered.

The two functions  $J_0(X)$  and  $(J_0(X) - J_2(X))$  are plotted in Fig. 16, and it can be seen that (1) they vary differently with  $X$ ; and (2) the ratio of the differences between their initial amplitude and their amplitudes at  $X$  is always close to the value of  $2/3$  derived above. For comparison, measured responses on a klystron driven to saturation are shown in Fig. 13.

## 8. Appendix 2. The Differential Phase Response

Information about this characteristic of the klystron can be obtained from the model of Appendix 1, by taking the amplitude compression coefficient,  $c$ , as zero but retaining the phase coefficient,  $k$ . It will be assumed that  $kP$  remains small enough for  $\cos(\omega t + kP)$  to be well

approximated by

$$\cos \omega t - kP \sin \omega t$$

Extension of the single-sideband work described in Appendix 1 then shows that, to first-order, the carrier is shifted in phase by  $k(P_c + 2P_m)$ , and the sideband by  $k(2P_c + P_m)$ . Still to first-order, no phase shift is introduced by gain compression, nor does a.m. to p.m. conversion have a direct effect on the gain of sidebands or carrier. The difference between the phase shifts at carrier and sideband frequencies, when  $P_m \ll P_c$ , indicates a source of differential phase distortion. The theory given here suggests that the observed value of differential phase shift, being the difference between carrier and sideband phase changes, is equal to the phase shift in the carrier alone, at a given carrier level.

While calculations of the level-dependent phase shift can be carried out for the ideal two-cavity klystron, it appears in practice that many other factors affect the value of differential phase shift recorded at a particular output level. For instance, one recent test has suggested that as the resonant frequency of the penultimate cavity is raised, the value of differential phase shift at the specified output level decreases; it is not clear at present whether this is because the a.m. to p.m. conversion coefficient changes, or whether other effects are involved. It is also known that feedback effects can perturb the differential phase response, though the extent of these perturbations has been greatly reduced by recent improvements to klystrons.

*Manuscript first received by the Institution on 27th February 1970, in revised form on 6th January 1971 and in final form on 23rd March 1971. (Paper No. 1406/Com. 50.)*

# A New Instrument for the Measurement of Vertical Currents in the Ocean

By

D. C. WEBB, M.Sc.,†

D. L. DORSON, B.Sc.†

and

A. D. VOORHIS, Ph.D.†

*Reprinted from the Proceedings of the Conference on Electronic Engineering in Ocean Technology held at Swansea from 21st to 24th September, 1970.*

A neutrally buoyant instrument able to measure and record vertical flow, water temperature, and pressure at any oceanic depth for extended periods is described. Principles of operation, design of the measurement, recorder, recovery, acoustic telemetry, and relocation sub-systems are discussed. Special features of the electronic design are very low power consumption and circuit flexibility to adapt the instrument to a range of oceanic experiments.

## 1. Introduction

The study of vertical current in the ocean has not been of central interest to Oceanographers, and it is improbable that the instrument described will be widely or routinely used. However if its novel capabilities are exploited in carefully designed experiments, it is possible to obtain valuable and revealing observations which would be unobtainable or very difficult to obtain using any other technique.

The central engineering design problems result from the operation of the instrument as a neutrally buoyant device, or Swallow float,<sup>1</sup> at any depth and for extended periods. Even with the use of structural materials of both high strength and stiffness to weight ratios, each unit of payload must be carefully utilized. Each component should be lightweight and all circuits should have low power drain. The acoustic telemetry and relocation system should be lightweight and efficient.

## 2. Principles

The measurement is made from a neutrally buoyant platform or Swallow float. The structure of the instrument is not only strong enough to resist the ambient hydrostatic pressure, but also stiff enough to be less compressible than water. Suitably ballasted, it will come into equilibrium at a predictable depth with water of known density and temperature.<sup>1</sup>

The instrument, shown in Fig. 1, is designed with a large metacentric height and is a solid of revolution about its vertical axis. Eight inclined vanes project outward and any vertical translation of water past the instrument acts on the vanes and rotates the whole instrument. The response of such an instrument to vertical flow has been discussed in reference 2. The angular position with respect to the magnetic meridian is measured with an internal magnetic compass. The approximate range of measurement is a minimum speed of 0.02 cm/s or less, and a maximum speed of 20 cm/s. An order of magnitude improvement at either end of the range is possible with a suitably modified design.

Physical properties of the water flowing vertically past the instrument, such as temperature, are measured conventionally. An acoustic signalling system permits

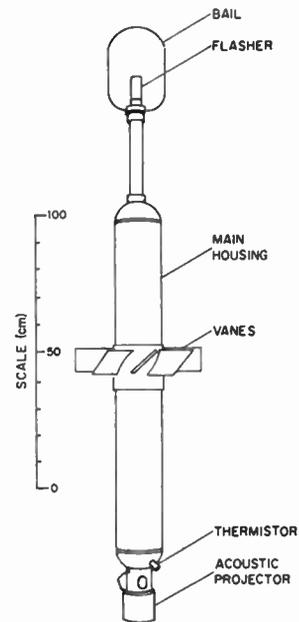


Fig. 1. Vertical current meter.

both relocation of the instrument and telemetry of some data. A weight is jettisoned and the instrument returns to the surface at the end of the experiment.

## 3. Experimental Applications

Observations relevant to several areas of ocean research can be made, each programme requiring careful consideration of the range, accuracy, and resolution of the variable to be measured.

Field programmes have been carried out in the following areas:

- (1) measurement of the vertical amplitude of internal waves;<sup>2</sup>
- (2) measurement of thermal convection in the ocean;<sup>3,4</sup>
- (3) measurement of thermal microstructure.

## 4. Mechanical Design

A straightforward mechanical design is used, the main housing is 6 in i.d.  $\times$   $\frac{1}{2}$  in wall (15.24 cm i.d.  $\times$  1.27 cm wall) aluminium commercial extruded tube, alloy 7075-T6, with hemispherical end caps of aluminium alloy A356-T6.

† Woods Hole Oceanographic Institution, Woods Hole, Mass. 02543, U.S.A.

The top bail is a recovery aid. A specially-designed release mechanism causes the ascent weight to be jet-tisoned. The end caps are held together by an internal tie rod. All sensors and the complete electronic package, with the exception of the flasher, are fastened to the lower end cap and are removable in one unit.

The 8 vanes are flat polypropylene square plates, 130 mm on a side and 10 mm thick. The vanes are mounted at an angle of 45° to the vertical, forcing the instrument to rotate once for every 110 cm of relative vertical flow. Both the vane assembly and the sonic transducer housing are welded polypropylene, a technique of fabrication that the authors have found very satisfactory for numerous underwater devices.

5. Electronic Design

The electronic equipment consists of three nearly autonomous sub-systems; acoustic relocation and telemetry, measurement and recording, and end-of-mission ballast jettison and flashing light. The sub-systems all use signals from a common clock and divider chain as well as a common power supply.

5.1 Acoustic Re-location and Telemetry

A sonic pulse of 5120 Hz carrier frequency, 8 ms duration and 40 W electrical input is transmitted every 4 s. The projector, a flooded ceramic cylinder (International Transducer Corporation Model 2003), is enclosed in a polypropylene housing and radiates approximately omnidirectionally. The signals are received by a rectangular array of hydrophones towed from a ship and after amplification, filtering, and demodulation, are recorded on a chemical recorder with a synchronized sweep and displayed on an oscilloscope.<sup>3</sup>

The maximum horizontal range at which signals are detectable depends on the local acoustic refraction; under good conditions ranges of 24 km are common.

One channel of data, usually pressure, is telemetered to the ship using pulse delay telemetry. A second acoustic pulse is transmitted after the main pulse, the delay being proportional to the variable being transmitted. This simple telemeter has been used extensively and successfully at the Woods Hole Oceanographic Institution.

5.1.1 Use of c.o.s.m.o.s. integrated circuits

The use of RCA complementary symmetry metal-oxide silicon integrated circuits (c.o.s.m.o.s.) with their very low power consumption (typically 10 nW for a binary toggle) permits a simpler, more integrated electronic design than would be possible with discrete components. Conventional micrologic devices consume enough power to exclude their use. The extensive use of c.o.s.m.o.s. devices in a wide range of equipment is probably the most important single development of the past several years in ocean instrumentation at the Woods Hole Oceanographic Institution.

5.1.2 Acoustic signal generator

The main divider chain is shown in detail in Fig. 2 to illustrate the use of c.o.s.m.o.s. devices and to show the source of many of the timing signals used throughout the instrument.

The operation of this circuit is self-explanatory. Note that, apart from the oscillator, only 6 integrated circuit packages are required. The oscillator is accurate to ±5 parts in 10<sup>6</sup> over the operating temperature range.

5.1.3 Acoustic telemetry

The circuit for generation of the data telemetry signal is shown in Fig. 3.

The individual resistances of the ladder networks have a value proportional to the binary value of the counter output. This is a conventional digital-to-analogue output circuit, and as the counter advances regularly, the output is the 128 step staircase waveform a. Note that a number of data channels can share this signal.

The c.o.s.m.o.s. devices, being complementary symmetry switches, are very suitable for this application, and the technique is applied several times in the instrument.

The CA3060 amplifier operates as a comparator and triggers the monostable when the ladder output a becomes more positive than the input signal.

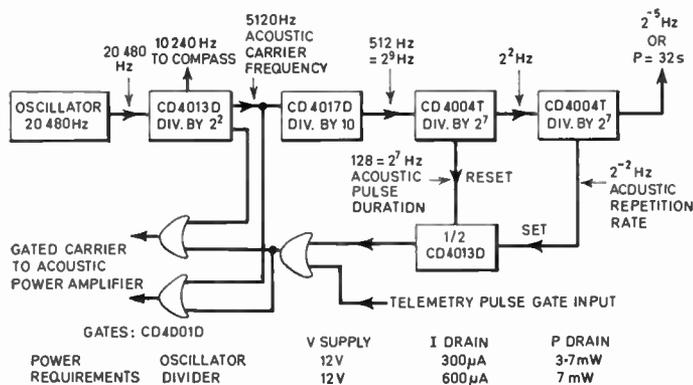


Fig. 2. Main divider chain.

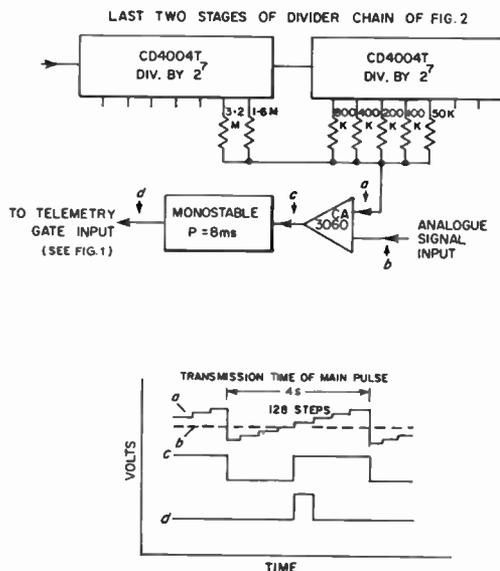


Fig. 3. Data telemetry circuit.

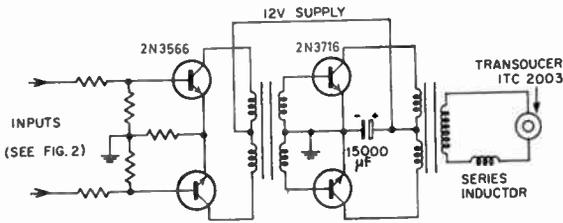


Fig. 4. Acoustic power amplifier.

5.1.4 Acoustic power amplifier

A conventional power amplifier using saturated switches and a series tuning inductor is shown in Fig. 4.

5.2 Measurement and Recording

The range of measurement and the recording arrangement are optimized for each experiment. A typical specification is as follows.

- PRESSURE: 30 to 70 bars—one traverse of recorder; resolution:  $\pm 0.4$  bars
- TEMPERATURE: 13° to 18°C two scales; coarse 5° per traverse, fine  $\frac{5}{8}$ ° per traverse; resolution  $\pm 0.05$ °
- ANGULAR POSITION: 0 to 360°  $\pm 10$ °—one traverse of recorder; resolution  $\pm 3$ °
- ENDURANCE: up to 60 days
- TIMING ACCURACY:  $\pm 10$  min.

5.2.1 Recorder

A strip chart recorder (modified Rustrak Model 88) is used. A current proportional to the various inputs is sequentially connected to the galvanometer stylus motor. The stylus is pressed against pressure sensitive paper once each second, and the paper is advanced continuously by d.c. chronometric motor (Escap AR 601 A1).

5.2.2 Pressure

A conventional pressure transducer, either potentiometer or strain gauge type, is connected to a low power operational amplifier. The amplifier output is suitable for operation of the recorder galvanometer and both zero and sensitivity are separately adjustable.

5.2.3 Temperature

Most experiments with the vertical current meter require unambiguous temperature measurements of

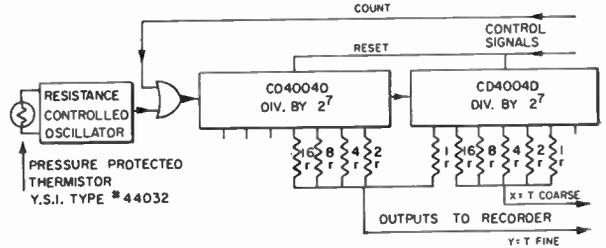


Fig. 5. Temperature measurement circuit.

medium accuracy and high resolution. Full scale deflection of the recorder for the full temperature range usually provides inadequate resolution, therefore the circuit of Fig. 5 is used. The counter circuit and digital to analogue ladder networks are similar to those of Figs. 2 and 3.

This technique has been used successfully in several applications and with input variables other than temperature. In the arrangement shown in Fig. 5, two digits of a number of base 16 are recorded, i.e. both x and y output have 16 possible values and the total value is  $16x + y$ .

Any convenient base (no. of resistors in the ladder) and number of digits (no. of ladders) can easily be selected for the special needs of any experiment, permitting data of arbitrarily high accuracy and resolution to be handled by a recorder or telemeter of comparatively poor performance.

5.2.4 Angular position

A suitable magnetic compass of moderate accuracy, low power consumption, high reliability and reasonable cost, is not readily available. The authors have been fortunate in obtaining a number of compass units from the U.S. Navy SSQ41A Sonobuoy. These compasses have conventional bearings and magnets; however the card or rotor is capacitively coupled to four stator electrodes forming a goniometer.

The four stator electrodes are excited with suitable signals of equal amplitude and quadrature phase, the magnetic angle being proportional to the phase of the signal coupled to the card. Figure 6 shows the circuit used.

The approach is slightly unconventional. The compass is excited with square waveforms, the direct output of

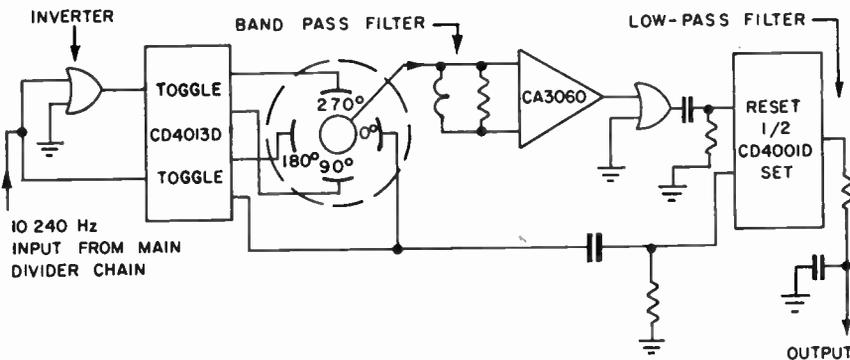


Fig. 6. Compass schematic.

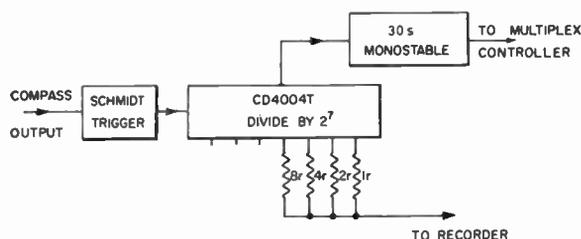


Fig. 7. Rotation counter.

c.o.s.m.o.s. toggles, assuring excitation of equal amplitude and phase quadrature. The bandpass filter on the output lead passes only the fundamental component of the excitation signal, and the set-reset flip-flop plus low-pass filter convert the phase information to a signal proportional to angle.

In certain experiments the instrument may rotate too fast for a coherent record of each rotation. In these cases, the circuit of Fig. 7 is used.

The counter advances each time the compass passes north in either direction. The counter has a capacity of  $2^7$  turns with a resistance ladder network connected to the last  $2^4$  stages.

Each time the least significant bit of the ladder network changes, i.e. every 8 turns, the new output is recorded. It is possible to use a unidirectional counter since the instrument always slows down to a speed where the direct recording of angle is readable before a reversal of direction.

### 5.2.5 Multiplexer

The various input signals are sequentially connected to the  $200\mu\text{A}$  movement of the recorder via a c.o.s.m.o.s. CD4016D multiplexer. The multiplexer control signals are derived from the main divider chain of Fig. 2.

The recording duty cycle of any variable is selected according to its importance and probable rate of change. The duty cycle also identifies the variable being recorded. One input, typically coarse temperature, is recorded for  $2^8\text{ s}$  ( $\sim 8.5\text{ min}$ ) every  $2^{12}\text{ s}$  ( $\sim 68\text{ min}$ ) to provide accurate time markers through the duration of the record.

### 5.3 End of Mission Recovery

It is the authors' goal that the measurement be terminated by an acoustic command from the recovery ship; however all instruments used to date have been recovered at a pre-set time.

A pulse of  $2^{-5}\text{ Hz}$  or 32 s period and approximately 0.1 s duration advances a Hecon electro-mechanical counter, Model FAO43. When the counter reaches a pre-set number, a switch closes and causes the ignition of a small propellant cartridge in a lightweight ballast jettison assembly<sup>5</sup> initiating the ascent of the instrument, and starting the flashing light.

The light is an O.A.R. Model SF500, which gives one flash of 0.1 joules each second, and is intended to supplement the acoustic system for relocation at the surface, particularly at short ranges at night.

### 5.4 Power Supply

The power requirements of the instrument are as follows:

acoustic relocation and telemetry	80 mW
recorder paper drive	40 mW
electronic circuits	10 mW
Total 130 mW	

The total energy required for 50-day missions is approximately 150 Wh.

Two cell types have been used successfully, Eveready Y 1471, an alkaline-manganese-zinc system, and Burgess RM 2550R, a mercuric oxide system with a special high surface area anode designed for operation at low temperature.

### 6. Analysis and Results

The record recovered at the end of an experiment is photographed and slightly enlarged. The data points are read from the photograph with dividers, or where critical with a travelling microscope, and the readings are transferred to punched cards.

Cards containing the data points and calibration curves for each instrument are fed to a computer which plots large, corrected curves of the variables in scientific units against time. These curves are the point of departure for the detailed scientific analysis of the experimental data.

An example of the data obtainable is shown in Fig. 8 and shows the large vertical water motions observable in a 'submarine cloud' in the North Western Mediterranean in winter.

### 7. Conclusions

It is possible to measure the vertical displacement, speed and physical variables of water flowing past a neutrally buoyant instrument. These data are recorded internally and the complete instrument returns to the surface for recovery at the end of the experiment.

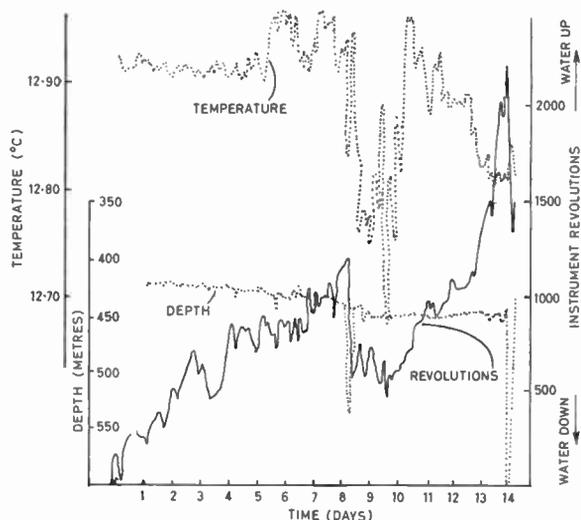


Fig. 8. Typical data for North Western Mediterranean.

A flexible, low-power electronic design permits the application of an instrument of moderate size in several areas of scientific research.

Future developments will probably be acoustic command recovery, and improved recording techniques, both analogue and digital.

### 8. Acknowledgments

Support of the U.S. Navy, Office of Naval Research Contract N000-66-C-0241; NR 083-004 is gratefully acknowledged as is the contribution of K. D. Fairhurst to the mechanical design. The paper is Contribution No. 2518 of the Woods Hole Oceanographic Institution.

### 9. References

1. Swallow, J. C. 'A neutral-buoyancy float for measuring deep currents', *Deep-Sea Res.*, 3, pp. 74-81, 1955.
2. Voorhis, A. D. 'Measurements of vertical motion and the partition of energy in the New England slope water', *Deep-Sea Res.*, 15, pp. 599-608, 1968.
3. Webb, D. C. and Worthington, L. V. 'Measurements of vertical water movement in the Cayman Basin', *Deep-Sea Res.*, 15, pp. 609-612, 1968.
4. Voorhis, A. D. and Webb, D. C. 'Large vertical currents observed in a winter sinking region of the North Western Mediterranean', *Cahiers Oceanographique*, June 1970.
5. Williams, A. J., III, and Fairhurst, K. D., 'Compact Ballast-Weight Release'. Woods Hole Oceanographic Institution Blue Cover Report. In preparation.

*Manuscript received by the Institution on 29th June 1970 (Paper No. 1407/IC50).*

© The Institution of Electronic and Radio Engineers, 1971

## The Authors



Mr. D. C. Webb obtained his B.Sc. degree from the Queen's University of Belfast in 1952 and his M.Sc. from Manchester University in 1954. He spent two years with Ferranti in Manchester and a further six years with Olivetti in Italy before joining Woods Hole Oceanographic Institution in 1962 where he is now a research specialist; Mr. Webb is a member of corporation of Bermuda Biological Station

for Research. His main research work is in the development of new observation techniques and he is author or co-author of 11 technical publications.



Dr. A. D. Voorhis holds the B.A. degree of Cornell University and the Ph.D. of Yale University. After five years' research in low temperature physics and neutron diffusion he took up an appointment at the Woods Hole Oceanographic Institution and since 1963 has been an associate scientist. He is author or co-author of 12 scientific publications in physical oceanography and author of about ten technical reports in

underwater sound and neutron diffusion problem. Dr. Voorhis's specialities in oceanography are hydrodynamics—particularly internal waves and vertical flow in the ocean.



Mr. D. L. Dorson graduated from Hiram College, Ohio, in 1965 and then joined the General Electric Company, Cleveland, Ohio, as a research physicist. In 1966 he went to Woods Hole Oceanographic Institution as a research assistant and since 1968 has been a research associate engaged on research and development of underwater acoustic equipment for instrumentation and communication.

# Use of an On-Line Digital Computer for Enhancement and Integration of Nuclear Resonance Spectra

By

**R. D. B. WAYMARK,**

B.Sc., (Graduate)†

and

**D. R. BOWMAN,**

C.Eng., M.I.E.R.E., A.M.B.I.M.†

*Reprinted from the Proceedings of the Conference on Laboratory Automation held in London from 10th to 12th November 1970.*

The paper describes the design and practical considerations involved in coupling a small digital computer to work on-line with a chemical research n.m.r. spectrometer. The various requirements of spectrometer, computer, interfacing and programming are dealt with both in general terms and in relation to the specific system employed in the present project. The effectiveness of the practical system is assessed, and some typical results are included.

## 1. Introduction

### 1.1 The Need for Enhancement

There has in recent years been a very substantial increase in the use of spectroscopic and allied equipment—all heavily dependent on electronic instrumentation—to assist in the analysis of substances in chemical, biological and medical studies.

At the same time there has been a growing demand for greater sensitivity in analytical instruments, that is, the capacity to produce from a weak sample a spectral output signal which is discernible from background noise signals.

Such background noise will be introduced primarily by the radiation source and associated detecting and amplifying circuits. However, once these noise contributions have been minimized, any further improvement in sensitivity (i.e. increase in signal/noise ratio) can only

be achieved by attention to the way in which the system is operated, rather than the design of its component parts. The process of improving sensitivity in this way is known as enhancement.

### 1.2 The C.A.T. Method

One well-established method of enhancement relies on the fact that true spectral signals will be coherent, that is, consistent in amplitude and position for repeated sweeps of the same spectrum. Background noise signals, on the other hand, will in general be completely random.

Hence if a digital computer is used to sum the results of repeated sweeps of a particular spectrum, any coherent signals will accumulate in direct proportion to the number of sweeps. Random noise signals, however, will tend to average each other out, and so accumulate more slowly. The averaging process gives rise to the term computer-of-average-transients (c.a.t.) to describe this type of enhancement technique.

A detailed statistical analysis of the c.a.t. method‡ shows that in an ideal case the coherent signal accumulates  $\sqrt{N}$  times faster than the random noise,  $N$  being the number of sweeps performed. This figure can be closely approached in practice, and so with a suitable  $N$ , a sensitivity improvement of at least an order of magnitude may be realized.

## 2. General Requirements of a C.A.T. System

The simplified block diagram of a typical c.a.t. system is shown in Fig. 1 and the function and requirements of the various parts will now be discussed.

### 2.1 Hardware Requirements

#### 2.1.1 Spectrometer

The spectrometer will, of course, supply the basic signal for enhancement and so will be required to cover the spectral range of interest for a given sample. Other requirements are as follows:

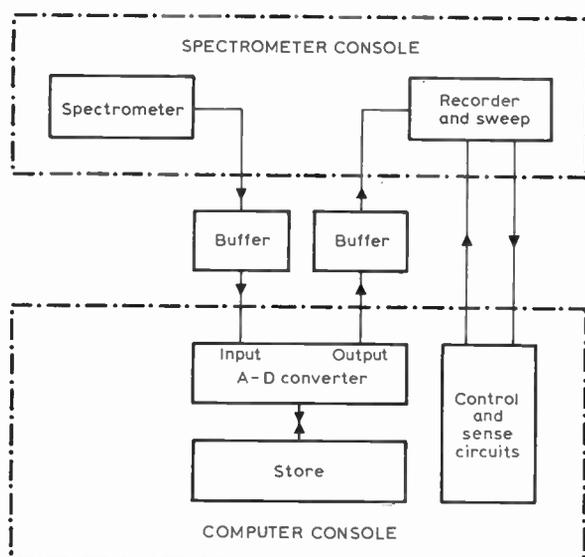


Fig. 1. Basic block diagram of a c.a.t. system.

† Department of Chemistry, University College London, Gower Street, London WC1.

‡ Ernst, R. R., 'Sensitivity enhancement in magnetic resonance. 1. Analysis of the method of time averaging', *Rev. Sci. Instrum.*, **36**, No. 12, pp. 1689-95, December 1965.

- (i) *Sweep.* The spectrometer must be capable of automatically performing repeated sweeps of the range of interest. It is important that the linearity of the sweeps is consistent, such that they can be perfectly superimposed. Some variation in the actual starting points of the sweeps may be tolerated as long as a suitable signal is available from a reference sample to act as a marker point on the spectral axis.
- (ii) *Recorder.* A chart recorder will often be incorporated as part of the spectrometer console, and generally the spectral sweep will be linked to the movement of the chart. A desirable feature is that the recorder should be readily disconnected from the spectrometer output, and connected instead to the digital-to-analogue output of the computer. This will make it possible to draw out enhanced signals with the same format as normal ones.
- (iii) *Stability.* The stability of the spectrometer with time will be an important consideration, since to perform a large number of sweeps will entail measurements over an extended period. Now the effect of drift in the spectral axis can be overcome if necessary by using a marker signal as noted under (i). However, such properties of the measured spectral signal as amplitude and line-width must be stabilized as much as possible in order to obtain maximum benefit from the enhancement process.

### 2.1.2 Computer

The primary function of the computer is to time-sample, accumulate and store the spectral signal during the series of sweeps comprising the accumulation process; then to read out these stored values to give the enhanced signal.

In addition, the computer is required to execute overall control throughout the enhancement and (if used) integration processes. For this it must be capable of sensing and acting upon such events as the start and finish of sweeps, and the completion of the required number of sweeps in an accumulation.

The requirements of a computer to fulfil these functions are tabulated below:

- (i) *Store.* The type of store used will to a large degree determine the overall operating speed of the computer. Again, the operating speed required for c.a.t. work will depend on the precise nature of the system involved. Thus for enhancement using sweeps of short duration, requiring a high sampling rate, a high-speed store (e.g. core or film) will be essential. On the other hand, if sweep times are longer, a low-speed store (e.g. magnetic drum) will be adequate.

Whatever type of store is chosen, it must have a capacity sufficient to accommodate the required sweep duration with time-sampling at a rate which achieves the desired resolution. Also a non-destructive read-out facility must be available.

- (ii) *Analogue-to-digital/digital-to-analogue converter.* This section of the computer changes analogue time-samples (i.e. voltages corresponding to the

spectral output at each of the sample points) to digital values for accumulation and storage. The converse function is performed when printing out the enhanced spectrum.

The two important requirements of the converter are, first, that conversion time is sufficiently rapid to be compatible with the overall sampling rate; and, secondly, that the dynamic range of the converter is such that the spectral input signals may be accommodated with acceptable resolution in the amplitude axis.

- (iii) *Sense and control facilities.* The computer must have facilities for interpreting external signals which give information about events occurring in the spectrometer. In particular, sense circuits must be available to respond to trigger signals sent by the spectrometer at the start and finish of sweeps. The computer must also incorporate control circuits which, when activated by the appropriate instruction, will initiate certain functions in the spectrometer (e.g. start the recorder sweeping, lower the pen to draw out a spectrum, etc.).

### 2.1.3 Interfacing

The term 'interfacing' is used to denote those facilities and connexions necessary in and between the spectrometer and computer so as to allow the passage of all necessary information between the two. The requirements of interfacing are tabulated under two headings:

- (i) *Signal circuits.* The output from the spectrometer must be matched in voltage amplitude and level to suit the computer a.-d. converter input. In practice this condition may not be met, and a buffer stage, consisting typically of a d.c. coupled low-frequency amplifier with backing-off facility, must be interposed between the two.

Again, the output of the a.-d. converter may not suit the recorder, and so another buffer stage will be required. As the available voltage will often be too high, a simple passive potential divider will suffice for this purpose.

- (ii) *Sense and control circuits.* The computer must sense two information signals, corresponding to the start and finish of a sweep, which tell it when to start and to stop accumulating (or reading out from the store). The finish signal may be derived from a microswitch mounted suitably on the recorder mechanism. The start signal must be rigidly tied to the sweep cycle (Sect. 2.1.1(i)), and may be derived in a number of ways: if the spectrometer is stabilized against drift in the spectral axis, then the start signal can be obtained directly from the recorder (assuming that this is coupled to the sweep generator) in the same way as for the finish signal. If the stability of the spectral axis cannot be relied upon, then it will be necessary to derive the start trigger signal from a secondary source. This often takes the form of a stable reference sample placed together with the main sample, which will provide a strong signal at a well defined point on the spectral axis.

Lastly, the computer must be able to execute remote control over the spectrometer. This will require that functions in the spectrometer such as starting the recorder sweeping and raising and lowering the pen can be performed by simple remote switching. The provision of these facilities will often entail small modifications or additions to existing mechanisms and circuitry.

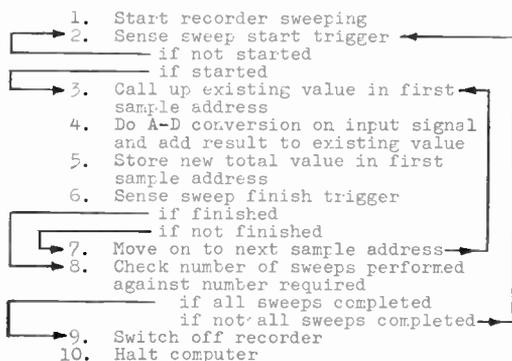
## 2.2 Software Requirements

This section deals with the design of a program for enhancement. The integration program is covered in a separate section.

### 2.2.1 Program for enhancement

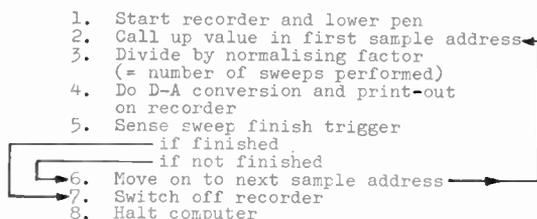
The enhancement program may be conveniently divided into two parts, namely accumulation and print-out. The essential steps in each are shown in the diagrams below:

#### (i) Accumulation



Steps 3 to 7 inclusive constitute the time-sampling cycle. The detailed program will have to be adjusted to give the required sampling-rate to achieve adequate resolution in a given experiment.

#### (ii) Print-out



### 2.2.2 General remarks about programming

Two further considerations are involved in the programming requirements for the c.a.t. system: first, the need for 'setting-up' programs which are run immediately prior to enhancement (and, if used, integration). These have the purpose of setting or clearing appropriate store positions and registers in preparation for running the main programs.

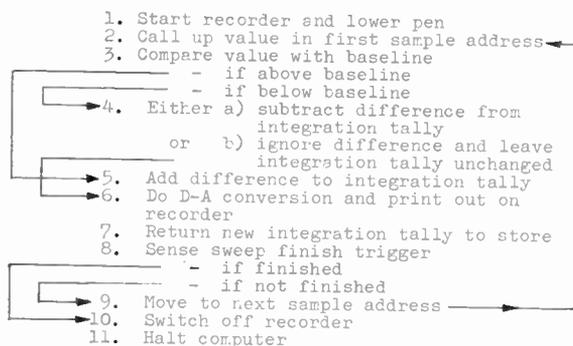
The second requirement, which has to be borne in mind when designing accumulation, print-out and integration programs is that it is highly desirable that all three

should operate on a compatible time scale. This will allow print-outs of enhanced and integrated spectra to be directly compared with each other, and the original un-enhanced spectrum.

## 3. Integration

This is the technique by which summation may be made of the area beneath the spectral peaks. Although a separate and distinct process from that of enhancement, integration by digital computer may be readily performed on material resulting from an enhancement experiment.

The essential steps in an integration program are shown below:



The alternative instruction (b) in step 4 will tend to give an integral with less 'wandering' and may be used to advantage with spectra of good signal/noise levels.

## 4. The Practical System

A block diagram of the practical system is given in Fig. 2. The various aspects of the system will now be described under the same headings as used in the general discussion of requirements in Sections 2 and 3.

### 4.1 Hardware Aspects of the Practical System

#### 4.1.1 Spectrometer

This is a Varian Associates HA100 high resolution n.m.r. spectrometer. It is primarily employed for proton ( $H_1$ ) resonance studies, with an operating frequency of 100 MHz and a magnetic field of approximately 23 kilogauss.

(i) and (ii) *Sweep and recorder.* The sweep system is mechanically linked to the recorder, which is of the flat-bed type. Either the applied frequency or the magnetic field may be swept in order to obtain an n.m.r. signal from the sample. Repetitive sweeping of a given range is controlled by micro-switches mounted at either end of the recorder carriage and associated circuitry (RL901 of Fig. 3). Sweep linearity is of a high order and enables the use of precalibrated chart paper.

A switching facility is available to isolate the recorder from the rest of the spectrometer and allow connexion to an external signal source.

(iii) *Stability.* Stability of the r.f. source in the spectrometer is readily secured by the use of a crystal-controlled oscillator. The magnetic field, on the

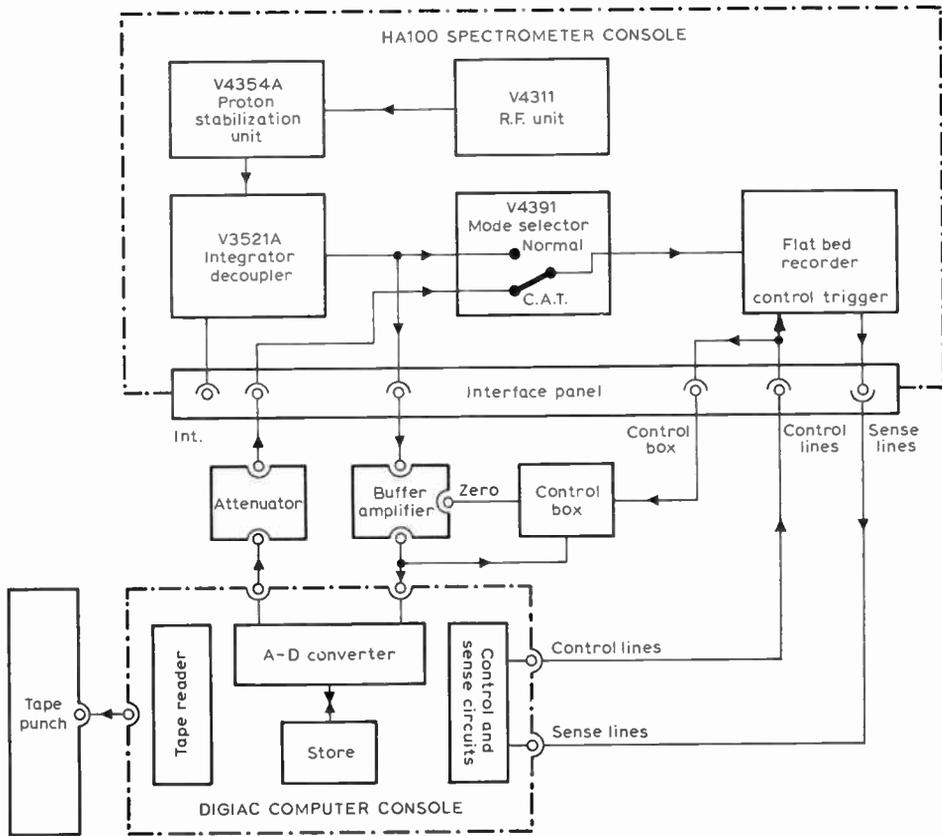


Fig. 2. Block diagram of the HA100/Digiac computer system.

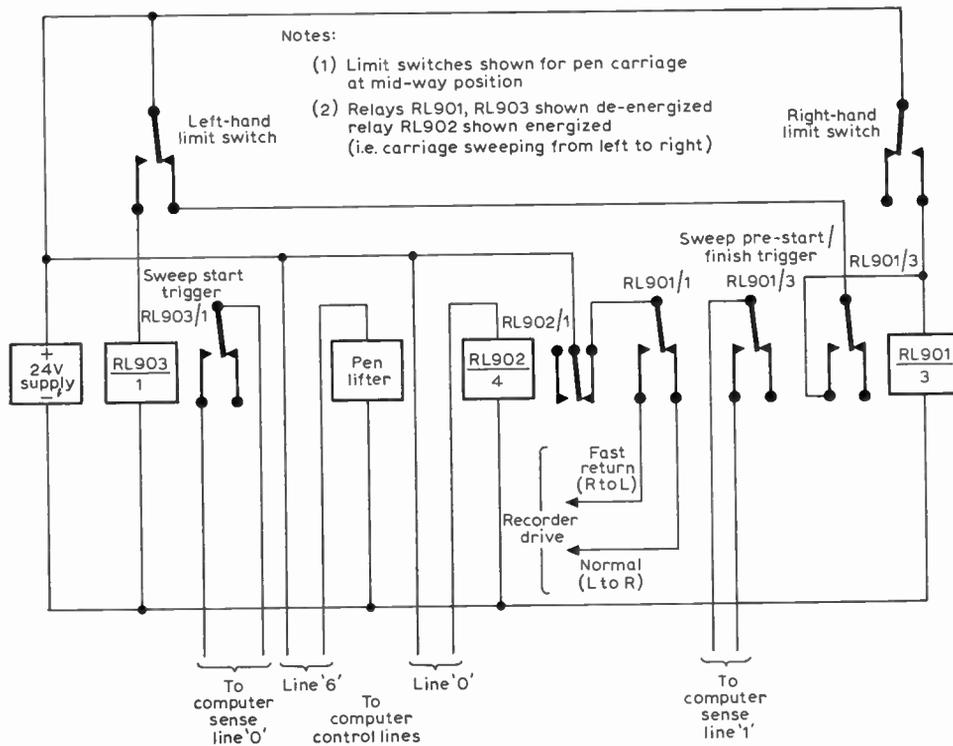


Fig. 3. Flat-bed recorder: drive and triggering circuits.

other hand, presents more of a problem, and several measures are necessary to reduce instability and drift to an acceptable level.

The magnet itself is a d.c. energized electro-magnet and is temperature stabilized by a thermostat controlled water flow. Electrical stabilization is then performed in two stages: the first employs sensing and compensating coils on the magnet. The second system, which achieves the extra stability required for high resolution work, senses the signal from a secondary (reference) sample placed together with the sample being investigated. Corrections are then made to the field such that the reference sample is always kept in a state of resonance.

Another property of the magnetic field is of importance in high resolution n.m.r. work. This is the uniformity, or homogeneity, of the field around the sample, which has a considerable effect on the amplitude and line-width of signals. In the HA100 spectrometer a system is available which automatically maintains the adjustment of homogeneity at optimum.

#### 4.1.2 Computer

The computer is a 'Digiac' low-cost general-purpose digital computer made by Digico Ltd. It is intended both for general computation and for on-line data collection, reduction and analysis.

- (i) *Store.* A magnetic drum store is used, with a random access time of 20 ms. The maximum sampling rate is therefore limited to about 20/s. This (or even a slower rate) is quite satisfactory for n.m.r. work, where comparatively slow sweep rates are employed.

The capacity of the store is 4096 16-bit words, of which up to 500 may typically be occupied by programs. Hence with a sampling rate of, say, 10/s, sweep durations of up to about 5 min can be accommodated.

Non-destructive read-out from the store is possible both when the computer is running and, by manual instruction, when it is halted.

- (ii) *Analogue-to-digital/digital-to-analogue converter.* This is an 8-bit successive comparison type with a 154  $\mu$ s. a.d. conversion time. The input range is approximately  $-1.5$  to  $+5.0$  V (see Fig. 4) giving a resolution capability in the amplitude axis of about 25 mV, corresponding to 1 part in 255.
- (iii) *Sense and control facilities.* The Digiac is equipped with eight sense lines which read in a '0' or a '1' depending on whether the lines are short circuit or open circuit with respect to a common line.

In addition, nine independent control lines are provided, these being connected to reed-relays which are activated by appropriate instructions from the computer.

Two further facilities are available in the Digiac computer: a tape-punch, with which tapes of both programs and spectra may be made, and a digital counter which may be used as a sweep tally during accumulation.

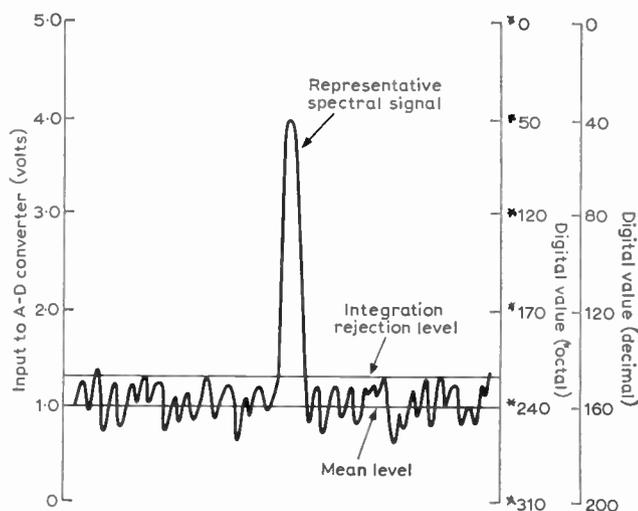


Fig. 4. Diagram showing action of analogue-to-digital converter.

#### 4.1.3 Interfacing

- (i) *Signal circuits.* The output of the spectrometer is a signal whose peak amplitude may reach a maximum of 300 mV, at a level slightly above chassis potential. To match this to the  $-1.5$  to  $+5$  V input requirement of the computer's a.-d. converter, the buffer amplifier of Fig. 5 is used. The circuit is built around the widely available '709' type linear integrated circuit; it has a voltage gain of about 30 from d.c. to 200 Hz and a backing-off range of  $-7.5$  to  $+7.5$  V in the output for a short circuited input.

The computer circuitry introduces an attenuation factor of approximately 3 times between a.-d. input and d.-a. output. The d.-a. output has then to be attenuated by a further factor of about 2 to suit the recorder, which has an f.s.d. of 1 V. This is performed by a simple 50 k $\Omega$  potential divider in the line between the d.-a. output and the recorder input.

A further facility in the practical system is the provision of a voltmeter to monitor the signal from the output of the buffer amplifier. The meter is mounted in the control box (Fig. 6) and serves as a useful guide when setting up. The control box also houses a remote zero setting control for the buffer amplifier.

- (ii) *Sense and control circuits.* The details of the sense and control circuits are shown in Figs. 3 and 6.

The sweep finish trigger (detected by sense line 1) is derived from RL901/3, the relay being activated by the right-hand limit switch on the recorder. The spectrometer is stable enough to allow the sweep start trigger also to be derived from the recorder, in the following way: when the carriage has returned to the left-hand side ready for a new sweep, RL901 is again activated, and RL901/3 now provides a 'pre-start trigger'. The sweep start trigger itself (detected by sense line 0) is then obtained from RL903/1 as the carriage moves off

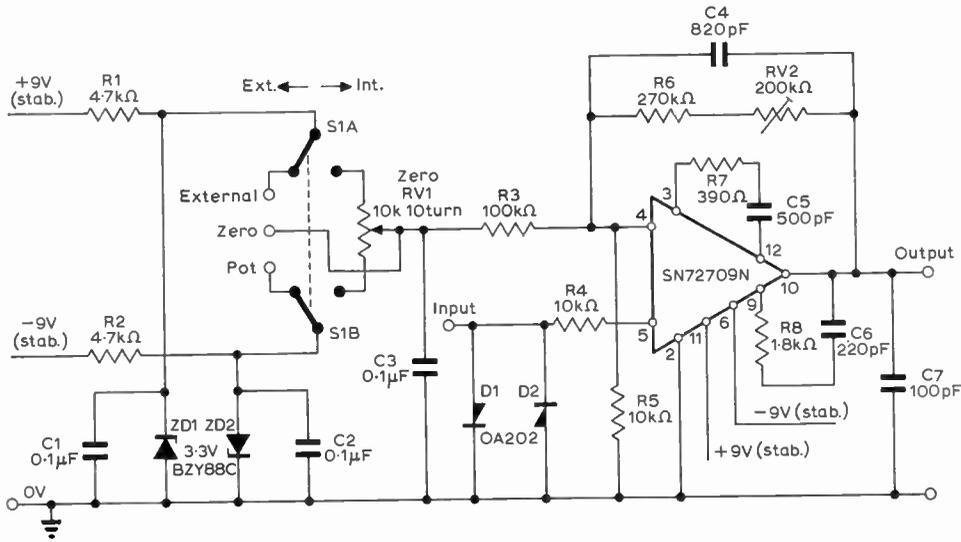


Fig. 5. Buffer amplifier: circuit diagram.

the left-hand limit switch. The reason for having two-stage start triggering is to avoid any uncertainty arising from variations in recorder 'turn-round time'.

As for the control circuits, control line 0 operates the recorder start and line 6 the pen lifter mechanism. Once started, the recorder will automatically perform repetitive sweeps until stopped again; similarly, when the pen lifter is activated it interlocks with the sweeps such that the pen is lowered during the sweep and raised during the sweep return.

Lines 1 to 5 are connected to indicator lamps mounted in the control-box and labelled as shown in Fig. 6. These are activated in turn at appropriate points during the running of a program and give a

visual indication that the program is operating correctly.

#### 4.2 Software Aspects of the Practical System

The practical programs for accumulation, print-out and integration will be considered in this section. They will be presented in outline only, and no Digiac machine language or instruction mnemonics included as these bear no direct relation to any of the widely used programming languages.

It is mentioned in passing that octal numbers (denoted by an asterisk, thus \*X) are used throughout in programming the computer. Octal notation is found to be the most workable compromise between binary notation, used in all internal computer operations, and decimal, the notation of everyday arithmetic.

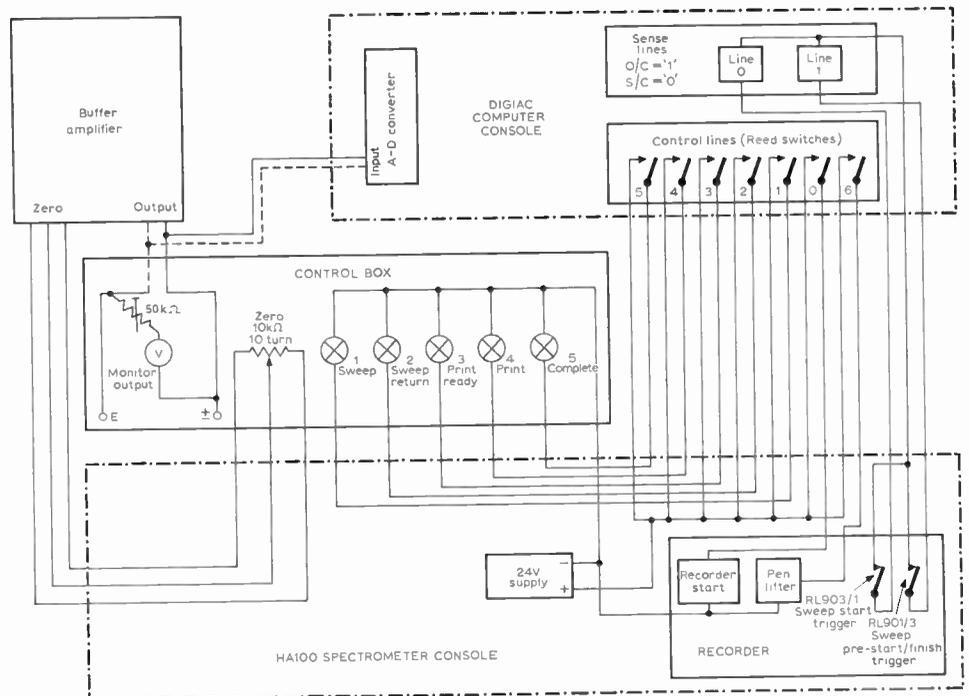
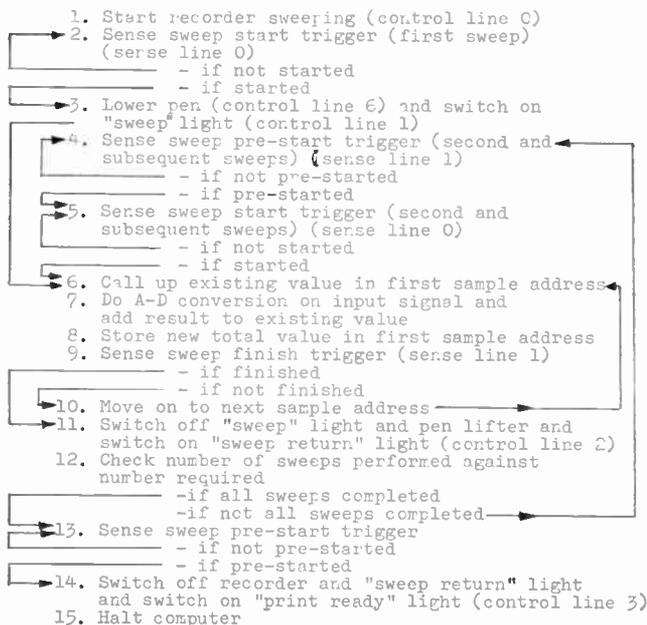


Fig. 6. Schematic showing control-box and interface wiring (sense and control lines only).

4.2.1 Practical program for enhancement

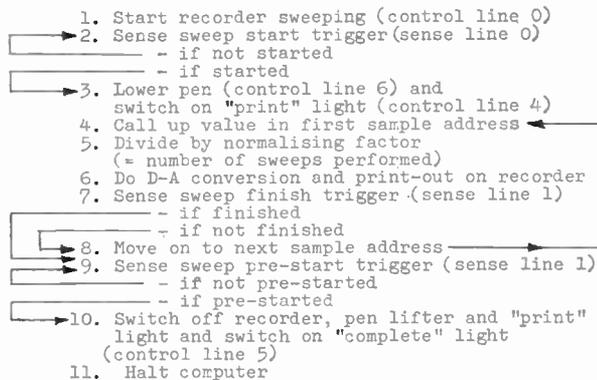
(i) Accumulation



Steps 6 to 10 inclusive, the time-sampling cycle, are designed in conjunction with the corresponding sections of the print-out and integration programs. By inserting 'padding' instructions where necessary, all three programs are able to run to a common time scale (cf. Sect. 2.2.2). The sampling-rate in the practical programs is about 8.5 samples/s which allows sweeps of up to 250 s to be accommodated with acceptable resolution.

A further point to note is that the pen lifter mechanism is activated during the first sweep only, such that only this sweep is drawn out. If traces of subsequent sweeps are also required, the pen lifter can be activated manually.

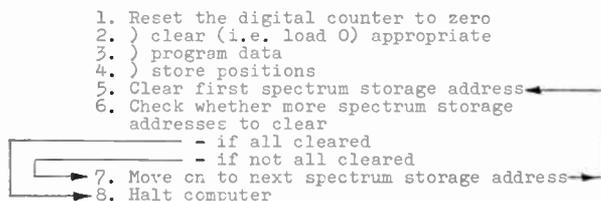
(ii) Print-out



The normalizing procedure (step 5) means that the enhanced spectrum will be printed out at the same vertical position on the chart as the input sweeps, without any need for adjustment of the recorder zero controls.

(iii) Pre-accumulation setting-up

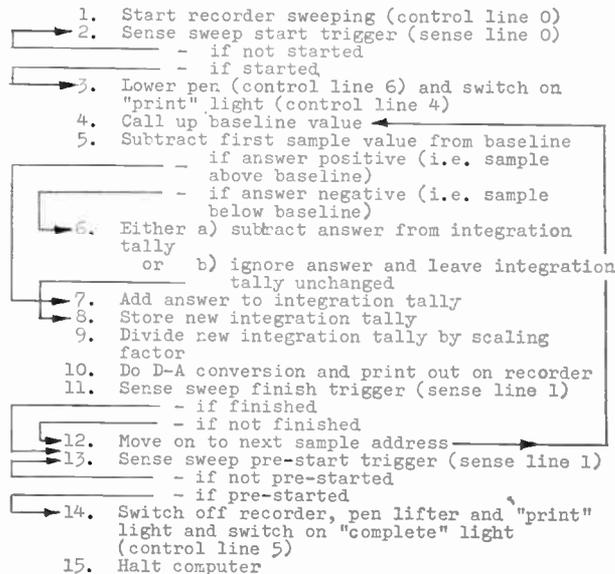
This program fulfils the functions discussed in Sect. 2.2.2.



Step 6 is included to prevent accidental erasure of program and program data information; in practice the program can either be left to run to completion, or stopped after the required number of spectrum storage positions have been cleared, when this is less than the total available number.

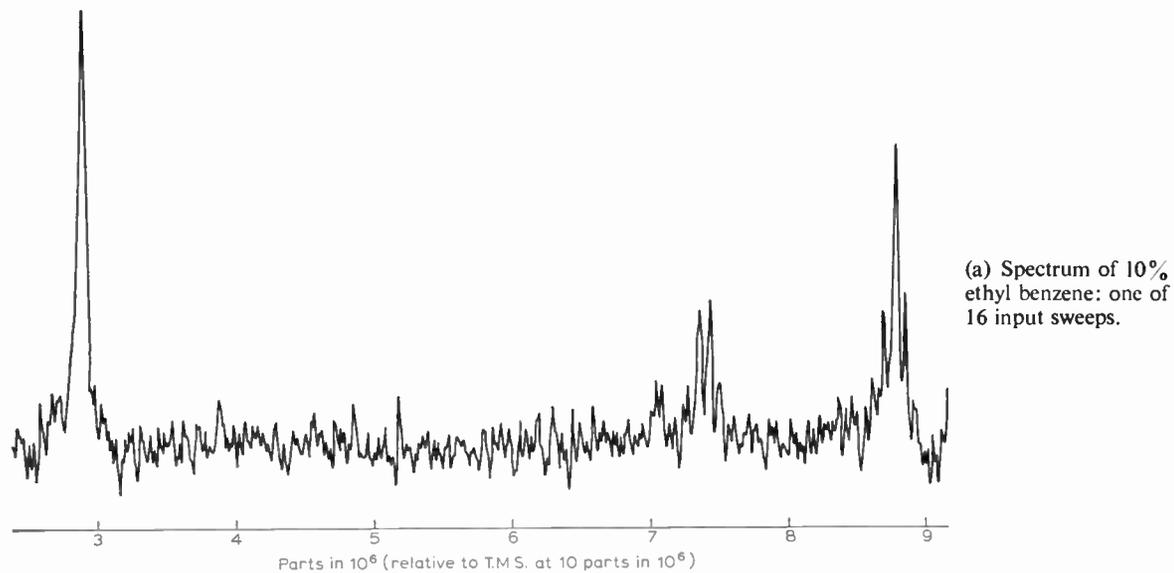
4.2.2 Practical program for integration

(i) Integration



It will be seen that a choice of instruction is given in step 6 (cf. step 4 in Sect. 3). Referring to Fig. 4, the baseline value chosen for program (i) corresponds to the line marked 'mean level', while that for program (ii) corresponds with the 'integration rejection level' line. Generally the two programs give similar results (Sect. 5.2), but there are instances in n.m.r. work where program (i) must be used in order to obtain valid results. These are the cases where a damped oscillation known as 'ringing' occurs about the baseline immediately following strong, sharp spectral lines. Clearly if program (ii) is used here, the positive ringing peaks will be integrated, while the negative peaks are ignored, resulting in an exaggerated value for the integral.

When program (i) is used it is necessary to assign a positive initial value to the integration tally (step 6a) to eliminate the possibility of the tally



(b) Spectrum of 10% ethyl benzene: enhanced output after 16 sweeps, with integral by program (ii).

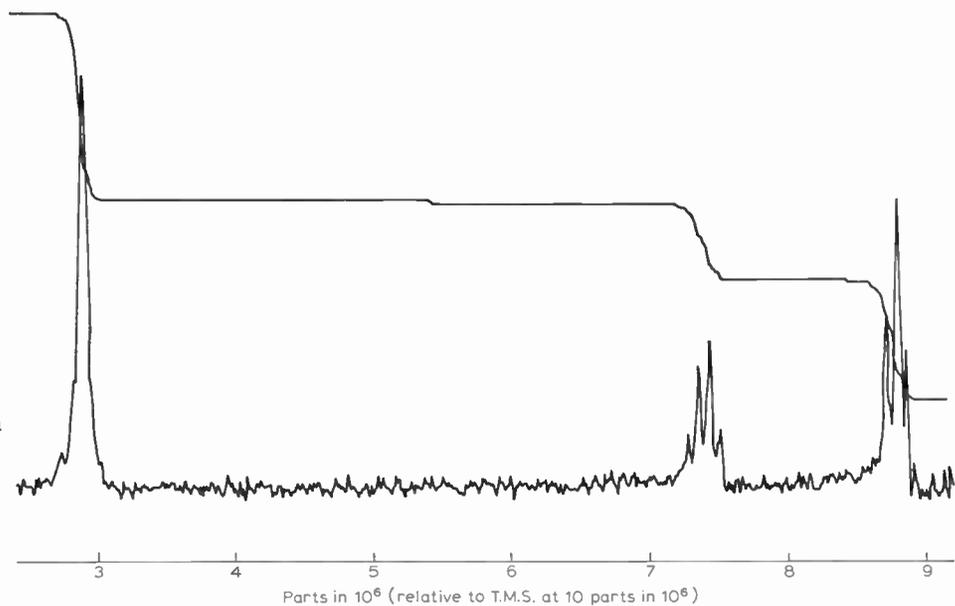


Fig. 7. N.m.r. spectra

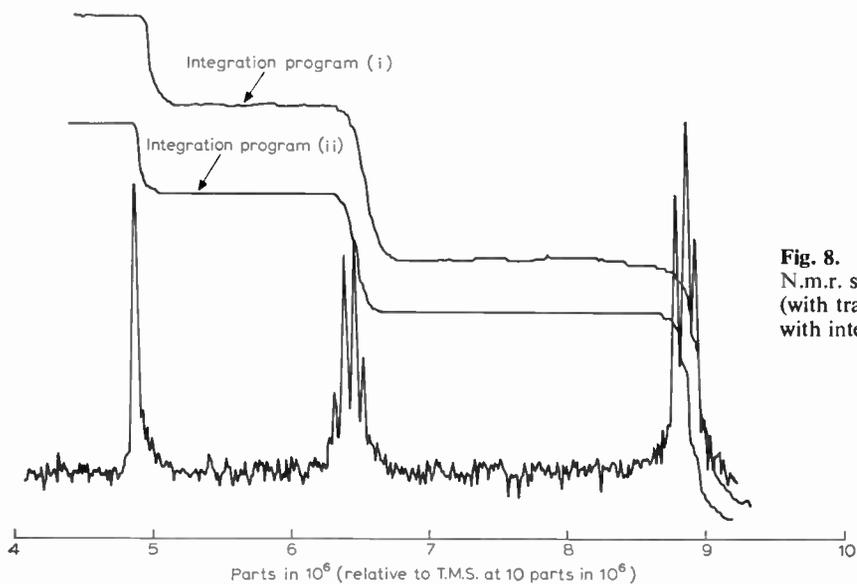


Fig. 8. N.m.r. spectrum of ethyl alcohol (with trace of HCl): one sweep, with integrals by programs (i) and (ii).

itself becoming negative if negative going peaks are encountered at the beginning of a spectrum.

The scaling factor (step 9) is included to ensure that values for d.-a. conversion always fall within the range of the d.-a. converter, i.e. \*0 to \*377. The value of the scaling factor necessary in a particular instance will depend on the numerical magnitude of the stored values comprising the spectrum, and therefore on the number of sweeps performed.

(ii) *Pre-integration setting-up.* This simple program is used before every integration run.

1. Clear (i.e. load 0) appropriate program data address
2. Load integration tally initial value into appropriate program data address
3. Halt computer

## 5. Evaluation of the Practical System

### 5.1 Assessment of Performance

At the time of writing the HA100/Digiac computer system has not yet been handed over for use as routine facility by research chemists. However, experience already gained indicates that the system is not only capable of producing valuable results but also is simple enough to be used by non-specialist operators. Some features of particular relevance in this connexion include:

- (i) *Ease of connexion of computer to spectrometer.* All connexions between computer and spectrometer are made on a single panel, and most may be left in permanently. When c.a.t. operation is required only one lead has to be connected and two switches set.
- (ii) *Simplicity of operation.* All program material is fed to the computer from pre-punched tapes. Minor alterations may be made manually—e.g. in the present program the number of sweeps in an accumulation may be set at any power of 2 up to 256. The need for other manual instructions is kept to a minimum by the use of setting-up programs and incorporation of setting-up instructions in the main programs. Once the computer is started, its activities and those of the spectrometer proceed to the completion of the program without any external attention.
- (iii) *Convenient presentation of results.* Positioning on the chart of enhanced spectrum print-outs is automatic once an initial run has been set up. Furthermore, enhanced spectra may be directly compared with normal runs since the same recorder and chart paper are used.

Copies of enhanced spectrum print-outs may readily be made; also spectra may be recorded on punched tape for reference or processing at a later stage.

### 5.2 Discussion of Results

Two specimen spectra (Figs. 7 and 8) are reproduced to demonstrate the effectiveness of the practical system in enhancement and integration.

(i) *Enhancement.* Figures 7(a) and 7(b) demonstrate the enhancement process in an accumulation

of 16 100s sweeps using a 10% ethyl benzene sample run deliberately under low sensitivity conditions. In Fig. 7(a) the peak signal to r.m.s. noise ratio is about 15:1. In Fig. 7(b), where the signal size is approximately maintained but the noise reduced, the enhanced signal/noise ratio is about 52:1. The improvement is thus a factor of 3.5, which closely approaches the expected value of 4 ( $= \sqrt{16}$ ) for 16 sweeps.

It is worth mentioning here the practical point that for all but the weakest samples there is little object in performing more than, say, 32 sweeps in an enhancement, since the improvement 'per extra sweep' above this value falls off very rapidly.

(ii) *Integration.* It should be emphasized that as far as the signal is concerned the accuracy of computer integration of n.m.r. spectra is subject to the same considerations as when conventional integration (by operational amplifier) is employed. Thus for valid results, such effects as line distortion due to incorrect phase adjustment or power saturation of the sample must be minimized.

Turning to the specimen results, Fig. 7(b) shows an integration performed on the enhanced spectrum of 10% ethyl benzene using program (ii). A valid integral is possible by this program since no ringing is present on any of the peaks. In fact the ratio of the number of protons in the three groups, corresponding to the singlet, quartet and triplet peaks is about 4.9 to 2.0 to 3.1 agreeing with the expected values of 5 to 2 to 3 to within 5%.

Figure 8 shows a single 100 s sweep of an ethyl alcohol sample (containing a trace of HCl) again run at lowered sensitivity. The signal/noise ratio is somewhat worse than for the enhanced ethyl benzene sample, and there is some pronounced distortion of the line shapes due to imperfect phase adjustment. Integrals by programs (i) and (ii) (Sect. 4.2.2) are plotted together, and, in the absence of any significant ringing effects, show substantially similar results.

## 6. Conclusions

This paper has described how the various design requirements of a c.a.t. system for spectrum enhancement have been met using a particular spectrometer and computer. Integration by computer has been included as an additional facility.

Experimental trials have shown that the practical system is both simple to use and effective in achieving enhancement figures close to the theoretical values. The integration programs have also been successful, although the need for care in the control of spectrum running conditions is stressed.

## 7. Acknowledgments

The authors would like to thank Dr. J. E. Anderson, Lecturer in Chemistry, for his interest in this work.

*Manuscript first received by the Institution on 28th August 1970 and in final form on 15th April 1971. (Paper No. 1408/Comp 135.)*

© The Institution of Electronic and Radio Engineers, 1971

# A Graphic Fixed Storage System for a Remote Terminal Display

By

**Professor G. R. HOFFMAN,**  
B.Sc., Ph.D.†

and

**J. K. BIRTWISTLE,** M.Sc.†

A series of linear elements are used to create a graphical display of a quality adequate for visual inspection. The graphic alpha-numeric character store described is a specially-constructed fast, pre-wired, closed magnetic path, transformer, read-only memory. However, the graphical techniques discussed could well be implemented by other storage methods. Experimental results and simulated displays are presented as an indication of the viability of the system. An automatic wiring machine has been developed to manufacture the store.

## 1. Introduction

Direct communication between an individual and a computing facility is a considerable advantage. In recent years several alpha-numeric display terminals have been developed for this purpose and may form part of the computing installation, or be remote and connected to a computer by landline. Because of the limited bandwidth available in this link line a remote terminal must also have local storage facilities. If low-grade storage methods are used to hold the data to be displayed, a high-grade store is then required for character generation.

This paper is concerned with a graphical generating system in which a series of linear elements (short straight lines of differing slope) are used to form graphical displays of a quality which is suitable for visual evaluation. These graphical 'characters' are generated in addition to a conventional range of alpha-numeric characters and both types are built up in the display by the scanning principle. The graphic-alpha-numeric character store described here is a fast, prewired, closed magnetic path, transformer, read-only memory.<sup>1,2,3</sup> A method of construction was adopted to enable the use of a numerically-controlled automatic machine in its manufacture. The wiring pattern of the store and the

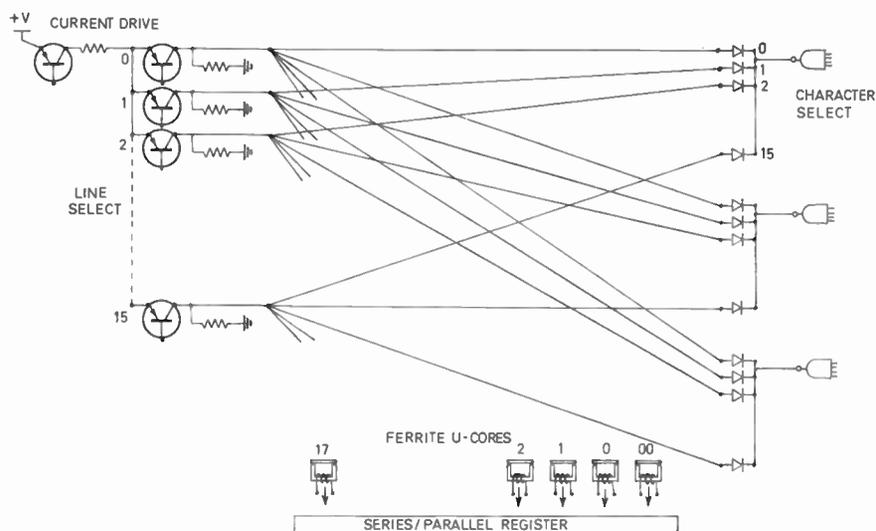
positioning of each wire termination are determined by information on a computer-prepared punched paper tape. This method of manufacture provides a storage medium which is particularly flexible and economic. It is acknowledged, however, that other storage methods<sup>4</sup> could also be used to implement the graphical technique described.

In order to provide a low-cost remote terminal it is convenient to use a television-type monitor as the display medium. If monitors of this type are utilized the e.h.t. and scanning circuits are substantially unaltered. The brightness modulation circuitry, however, must be modified extensively as the video bandwidth required for an alpha-numeric or graphical display is at least twice that needed for a television picture.

## 2. The Display System

### 2.1 The Character Store

The working of the display system is more easily understood if the character store (Fig. 1) is considered first, as its arrangement determines some other aspects of the system design. Each character has associated with it sixteen diodes and sixteen wires which thread or bypass each of a row of ferrite U-cores in a pattern which is related to the shape of displayed character.



**Fig. 1.**  
Schematic diagram  
of 'read-only' store.

† Electronic and Electrical Engineering Department, The University of Manchester, Oxford Road, Manchester 13.

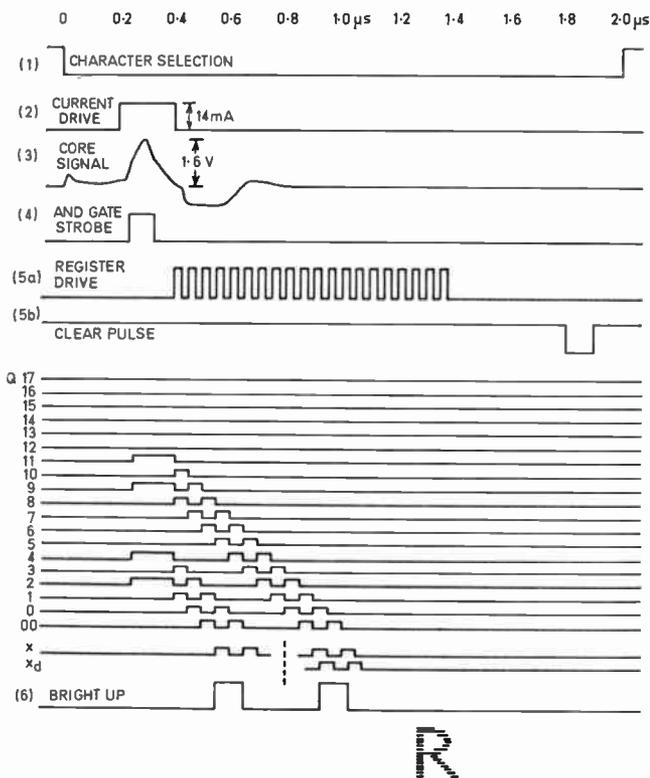


Fig. 2. Timing diagram of line 4 of character R.

- (1) Character selection pulse.
- (2) Current drive.
- (3) Transformer signal.
- (4) AND gate strobe.
- (5a) Register serial drive pulse train.
- (5b) Register clear pulse.
- Q<sub>17</sub>-Q<sub>00</sub> The output waveform of each of the flip-flops in the register.
- (x) Normal output.
- (x<sub>d</sub>) Delayed output.
- (6) Bright-up waveform.

To select a character an associated AND gate is switched to its low state and the wires which thread the core array are selected in sequence by the line select current drive transistors. A secondary winding on each ferrite core supplies signals from those transformers which are threaded by a selected wire to load a series-parallel register. Series operation of the register (an assembly of D type edge-triggered flip-flops) then provides brightness modulation for one line of the chosen character. Each bright-up pulse is initiated by the output signal obtained when a selected wire threads a transformer, and is terminated by a signal from the next transformer along the row to be threaded by the wire. It should be noted that when a row of characters is displayed the first line of each character along the row is selected in sequence before the second line of the leading character is started.

The timing sequence for the fourth line of the character R is considered in detail in Fig. 2. The character R is selected (1) and the current drive pulse initiated in the fourth line select transistor (2). Transformers which are threaded by this wire give rise to output signals (3) which, suitably strobed (4), actuate Q<sub>2</sub> Q<sub>4</sub> Q<sub>9</sub> and Q<sub>11</sub> of the

register, setting these flip-flops simultaneously into the 'one' state. The register now contains the information needed to generate line 4 of the character R. It is then driven serially with a high frequency waveform (5) which empties the register and a divide-by-two circuit at the output provides the modulation signal (6) required to display the selected line of the character.

In the storage system thus described two distinct operations occur: the information is first transferred in parallel from the transformer store into the register and then is discharged from the register in serial form to produce brightness modulation pulses for one line of a character. The register is refilled in the time interval used to provide spaces between characters.

The system arrangement avoids the time delay associated with the charging of the relatively high capacitance between a selected wire and the remaining wires. When a line is selected all the wires move together in potential, but only the wire which is terminated by the selected character AND gate carries line drive current. A more detailed consideration shows that the store is a complex array of coupled transmission lines which must be terminated at one end by their characteristic impedance to prevent undue ringing. Despite this requirement the signal from the linear ferrite transformer is available in less than 100 ns.

The horizontal resolution is primarily determined by the number of transformers in the array and may be doubled by including a 'logic' transformer. Wires which thread this transformer produce a signal which causes the bright-up pulse to be delayed by an interval corresponding to half the register drive pulse period. A second logic core allows the technique to be extended, enabling a second bright-up pulse to be shifted relative to the first. This is especially useful when processing letters such as W which have counter-sloping strokes; an effective horizontal resolution of 32 points (i.e. the ability to brighten or extinguish the display spot at any of 32 positions across the character width) is obtained. A selection of simulated alpha-numeric characters with an effective horizontal resolution of 32 points is shown in Fig. 3.



Fig. 3. Some alpha-numeric characters.

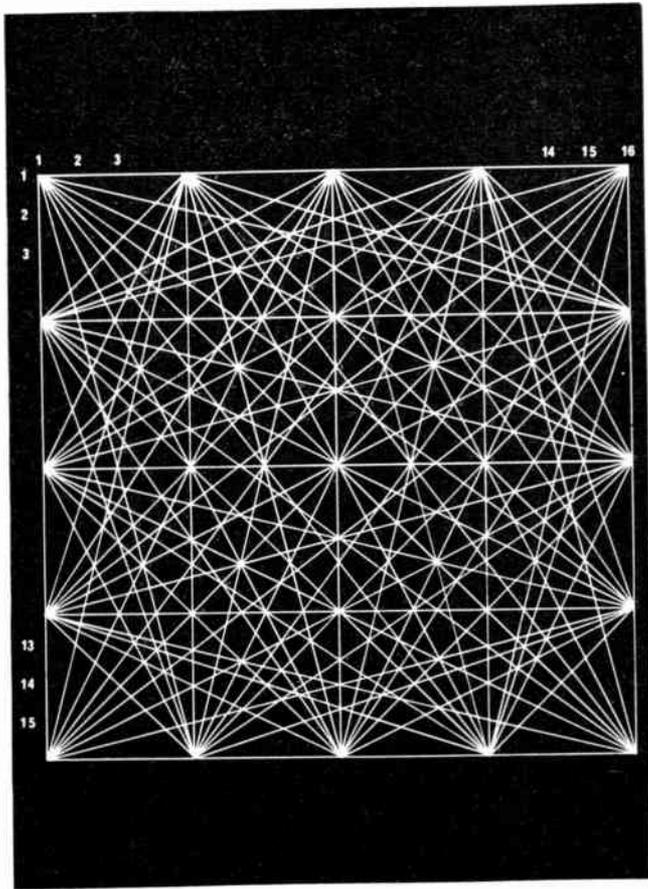


Fig. 4. A unit cell with all lines present.

## 2.2 Graphical Display

To obtain a graphical display the screen area is divided into square unit cells, each approximately the size of an alpha-numeric character. The edges of the unit cells are divided into four equal parts, forming sixteen points on the perimeter of each cell. If lines are drawn from each peripheral point to all of the others the total number (excluding redundancies) is 120. A unit cell with all the lines present is illustrated in Fig. 4. Using such a set of linear graphical 'characters' a series of short lines can be constructed over the display area to represent any desired graphical form. Some simulated displays, namely a damped oscillation and a large and small circle, are shown in Fig. 5. It should be observed that the linear graphical elements are also formed by the scanning technique used for alpha-numeric characters and a number of typical examples are shown inset in expanded form in the same figure. Because of the time interval required to load and discharge the series-parallel register, alternate unit cells in each row are selected in a scan and the remaining interleaving cells are displayed in the following scan.

### 2.2.1 Mirror imaging

In the graphical display, information about the character shape is contained in the pattern in which the fifteen line select wires are threaded through the transformer array. 1800 wires and diodes would be needed to store 120 graphical characters. However, a high degree of symmetry exists in the unit cell and a simple modification can reduce the requirement by a factor of four, a considerable reduction in the cost and complexity of the store.

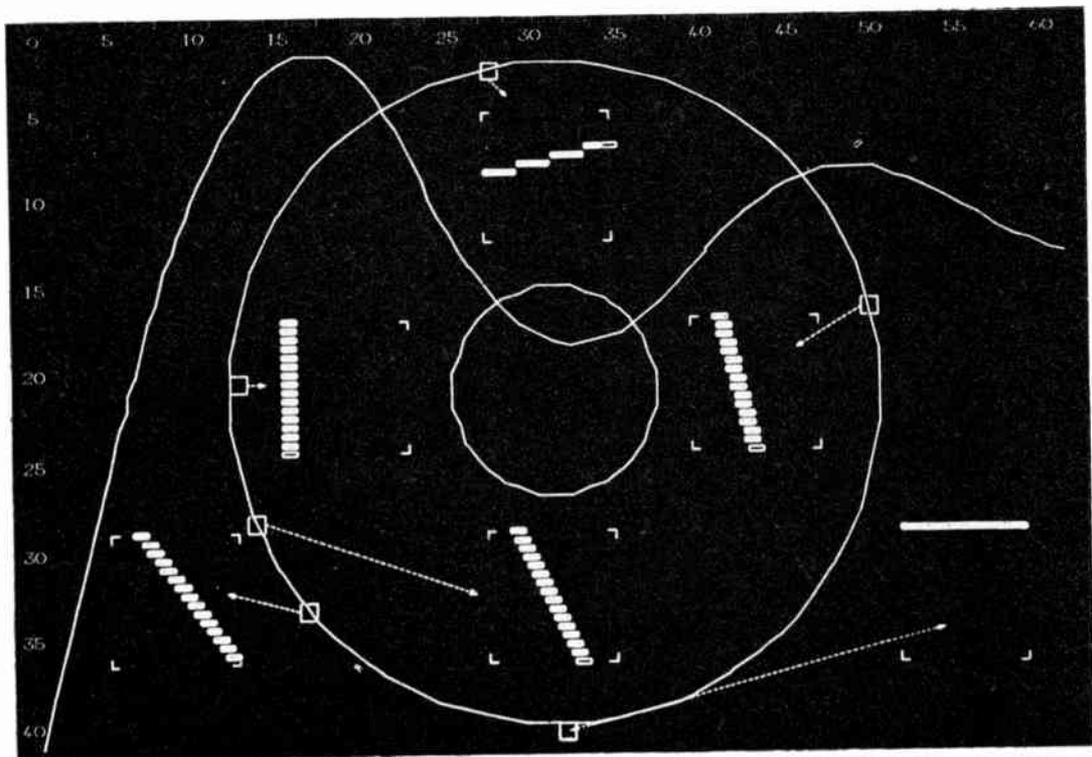


Fig. 5. A simulated graphic display.

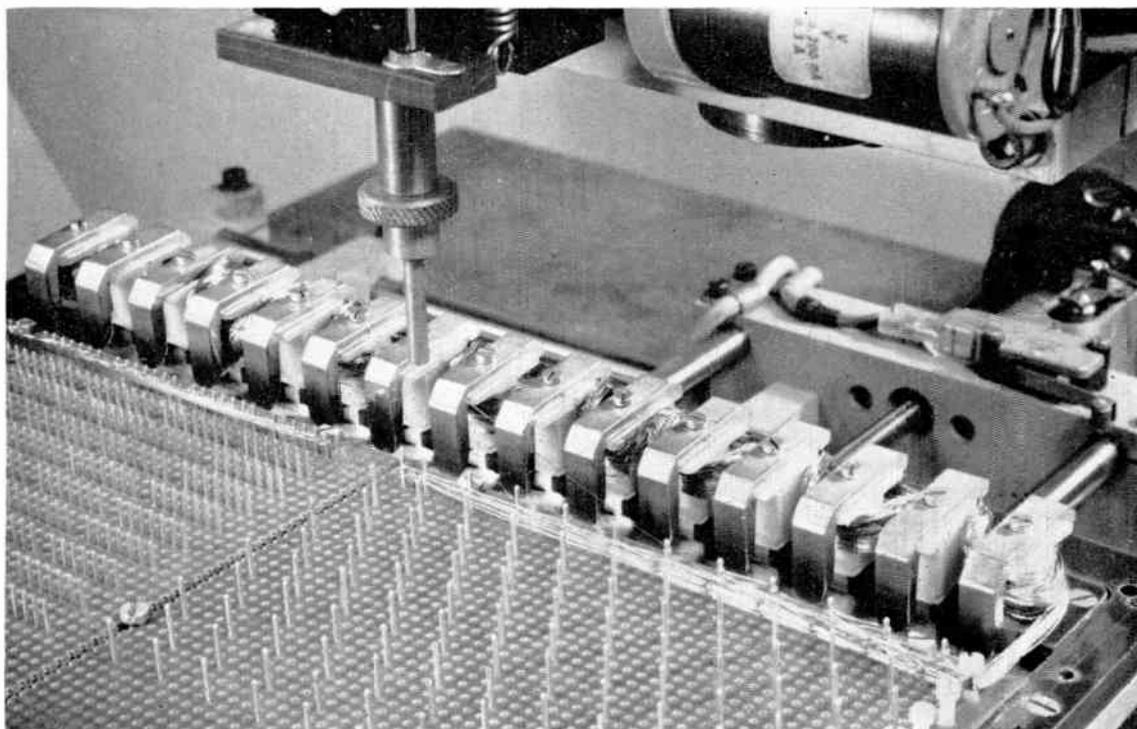


Fig. 6. A partially-wound store.

To obtain the vertical mirror image (lateral inversion) of a graphical character the discharge direction of the series-parallel register is reversed. Similarly reversal of the selection sequence of the line select drivers leads to a horizontal mirror image of the character. If the mirror image technique is adopted the number of graphical characters needed in the store is reduced to 32.

### 2.2.2 Overlap

Because the series-parallel register is an incomplete shift register, adjacent transformers in the array cannot be selected by wire threading. This restriction sets both the minimum bright-up pulse width and the minimum spacing between pulses. To preserve horizontal continuity between unit cells it is desirable to overlap at one edge and this is made possible by adding another transformer to the character store array. Vertical continuity between unit cells is maintained by using fifteen lines of a sixteen-line cell; the sixteenth line is the first line of the adjacent lower cell. Both of these effects are illustrated by pulses shown in outline in the inset diagrams of Fig. 5. The method of obtaining vertical continuity and the mirror imaging technique both require additional delay in the bright-up waveform when the line select circuitry is driven in opposite sequence. This delay is equal to one period of the register drive waveform.

## 3. The Automatic Wiring Machine

A typical repertoire of 128 alpha-numeric characters and 120 graphical characters with mirror imaging requires a character store having twenty transformers with 2250 wires and 4500 terminations. Manufacture of the store by normal manual methods would be a formidable task and automatic assembly was considered essential.

A machine capable of being positioned in three dimensions under punched paper tape control has been adapted to wire the character store. Two degrees of positional accuracy are employed. Normal machine instructions allow wires to be routed through the transformer array at the full working speed of the machine. Each wire is terminated by machine-wrapping on to an integrated circuit terminal pin. This operation requires precise, positive location of the wire-wrap tool above the terminal pin before the wrapping sequence is initiated and takes longer to complete than a normal command.

An individual line select wire is placed in the store in the following manner. Polyurethane-insulated copper wire is wrapped on a pin in the line select terminal plane, laid through the transformer array, threading the appropriate bobbins and wrapped on a pin in the character select terminal plane. Wire is then wrapped on an adjacent pin in the character select plane and returned through the transformers to the line select plane. This process, photographed in progress in Fig. 6, is continued until all the wires have been laid in the store. The short wire links remaining between adjacent terminal pins must be removed manually and the wire wrap connexions soldered to ensure reliable electrical contact.

In order to avoid these manual operations another technique is currently being investigated in which the wire connexions are made by reflow soldering. The insulated wire is held in contact with a solder- and flux-coated pad on the termination plane by a wedge-tipped probe. A short heat pulse at the probe tip causes breakdown of the insulation and allows solder to flow round the wire. Tensioning the wire with the probe in position causes it to break at a point close to the probe tip. When the free end of the wire is repositioned under the probe

tip the device is then ready to bond a new length of wire between another pair of terminal pads. The successful incorporation of this technique in the wiring machine would allow wiring of the character store to be completely automatic.

**4. Conclusions**

The number of characters which can be displayed on a television-type monitor screen is approximately 60 by 40, or 2400. If 128 alpha-numeric characters and 120 graphical characters are stored an eight-bit number is needed to specify a character. A total of 19,200 bits must also, therefore, be accommodated in a local store. More complex addressing separating graphic from alpha-numeric symbols could reduce this requirement. This is a relatively expensive part of the system and for this reason experimental work has been confined at present to the character store.

A variety of linear graphical elements and alpha-numeric characters have been generated and displayed on a c.r.t. screen: the character store has been filled with

wire and problems associated with large interacting systems have been simulated. It would appear that the method described for producing graphical information on an inexpensive display terminal using only 120 graphical elements is a feasible proposition.

**5. References**

1. Dimond, T. L., 'No. 5 crossbar AMA translator,' *Bell Laboratories Record*, p. 62, February 1951.
2. Knighton, B. W. and Taub, D. M. 'The design of transformer (Dimond ring) read only stores,' *I.B.M. J. Res. Development*, **8**, p. 443, September 1964.
3. Aldrich, W. H. and Alonso, R. L. 'The braid transformer memory,' *I.E.E.E. Transactions on Electronic Computers*, EC-15, No. 4, p. 502, August 1966.
4. Taub, D. M. 'A short review of read-only memories,' *Proc. Instn Elect. Engrs*, **110**, p. 157, 1963.

*Manuscript first received by the Institution on 2nd November 1970 and in final form on 29th March 1971. (Paper No. 1409/Comp. 136.)*

© The Institution of Electronic and Radio Engineers, 1971

**STANDARD FREQUENCY TRANSMISSIONS—August 1971**

(Communication from the National Physical Laboratory)

August 1971	Deviation from nominal frequency in parts in 10 <sup>10</sup> (24-hour mean centred on 0300 UT)			Relative phase readings in microseconds N.P.L.—Station (Readings at 1500 UT)		August 1971	Deviation from nominal frequency in parts in 10 <sup>10</sup> (24-hour mean centred on 0300 UT)			Relative phase readings in microseconds N.P.L.—Station (Readings at 1500 UT)	
	GBR 16 kHz	MSF 60 kHz	Droitwich 200 kHz	*GBR 16 kHz	†MSF 60kHz		GBR 16 kHz	MSF 60 kHz	Droitwich 200 kHz	*GBR 16 kHz	†MSF 60 kHz
1	-299.9	+0.2	+0.1	520	545.7	17	-300.0	0	+0.1	508	537.3
2	-299.8	+0.1	+0.2	518	544.5	18	-300.0	0	+0.1	508	537.0
3	-299.9	+0.2	+0.1	517	543.0	19	-300.0	-0.1	+0.2	508	537.9
4	-299.9	+0.1	+0.1	516	542.4	20	-300.0	0	+0.1	508	537.5
5	-299.9	+0.1	+0.1	515	541.5	21	-300.0	0	+0.1	508	537.4
6	-299.9	+0.1	+0.1	514	541.4	22	-300.0	0	+0.1	508	537.8
7	-299.9	+0.1	+0.1	513	540.3	23	-300.0	0	+0.2	508	537.6
8	-299.9	+0.1	+0.2	502	539.4	24	-299.9	+0.1	+0.2	507	537.0
9	-299.8	+0.2	+0.2	510	538.0	25	-300.0	0	+0.2	507	537.1
10	-299.9	+0.2	+0.1	509	540.5	26	-300.0	+0.1	+0.1	507	536.0
11	-300.0	+0.1	+0.1	509	538.6	27	-300.0	0	+0.1	507	536.0
12	-299.9	+0.1	+0.1	508	537.8	28	-299.9	0	+0.1	506	535.8
13	-300.0	0	+0.1	508	537.5	29	-299.9	+0.1	+0.1	505	535.0
14	-300.0	0	+0.1	508	537.7	30	-299.9	0	+0.1	504	534.6
15	-300.0	+0.1	+0.1	508	537.0	31	-300.0	+0.1	+0.1	504	535.5
16	-300.0	0	+0.2	508	536.9						

All measurements in terms of H.P. Caesium Standard No. 334, which agrees with the N.P.L. Caesium Standard to 1 part in 10<sup>11</sup>.

\* Relative to UTC Scale; (UTC<sub>NPL</sub> - Station) = + 500 at 1500 UT 31st December 1968.

† Relative to AT Scale; (AT<sub>NPL</sub> - Station) = + 468.6 at 1500 UT 31st December 1968.

# Transatlantic Telephone Cable between U.K. and Canada

CANTAT 2, the new high-capacity U.K.-Canada submarine cable that will more than double telephone links across the Atlantic, will be made for the British Post Office and the Canadian Overseas Telecommunications Corporation by Standard Telephones and Cables Ltd. With 1840 circuits, the cable will have over 400 more circuits than all the existing transatlantic cables put together; and, in addition to telephone calls, it will be used for telex, telegrams and data transmission. It will be produced to an advanced design made possible by work at the Post Office Research Station at Dollis Hill, North London.

More than 2800 nautical miles of cable will be laid to cover the 2700-nautical-mile route between the British cable station in Widemouth Bay, Cornwall, and a new cable terminal to be built near Halifax, Nova Scotia. Altogether the complete cable will weigh little more than 15 000 tons, and for most of its length it will be less than two inches in diameter.

Bringing the cable into service will cost about £30.5 M. This covers production (amounting to £22 M), survey and development work and the laying operations. The cost will be shared by the British Post Office and the Canadian Overseas Telecommunications Corporation. Two cable ships will be chartered to lay the cable. They are the Cable and Wireless cable ship *Mercury* and the powerful Canadian ice-breaker/cable-laying ship *John Cabot*.

CANTAT 2 will be the third undersea cable between the United Kingdom and Canada. The first transatlantic telephone cable—TAT 1, from Oban, Scotland to Clarenville, Newfoundland—was opened in 1956. CANTAT 1, first section of the Commonwealth Cable Network was laid between Oban and Hampden, Newfoundland, in 1961, with 80 circuits to carry calls between Britain and Canada and on to New Zealand, Australia and the Far East over the Commonwealth Pacific and South East Asia cable systems.

Standard Telephones and Cables Ltd. of London have also been awarded an export contract worth £10 M to manufac-

ture and install a 160-circuit submarine telephone cable between Brazil and the Canary Islands. The cable system—code-named BRACAN I—will be Brazil's first direct cable link to Europe, and will supplement the existing satellite links. It will be ready for service at the end of 1972.

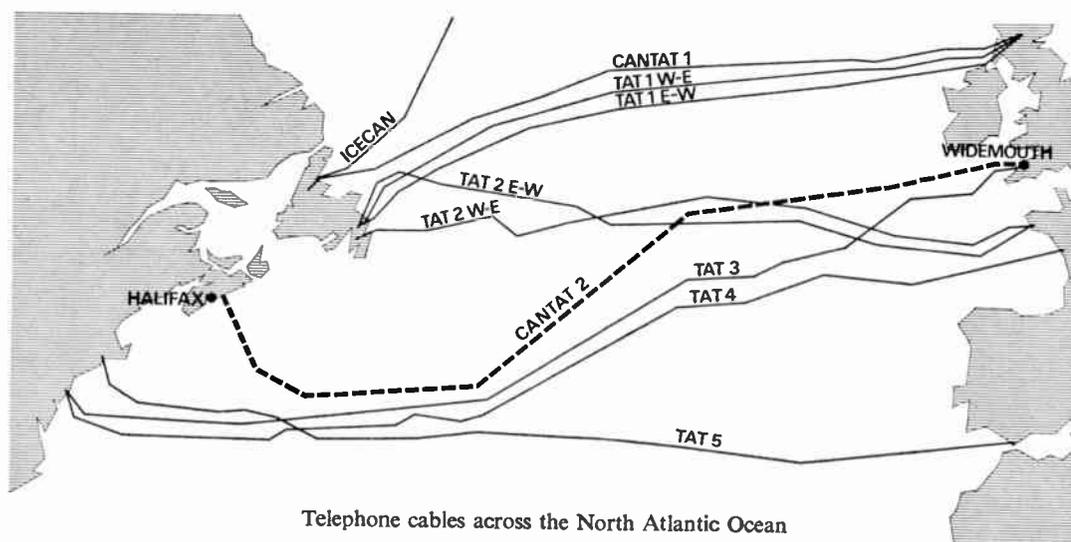
The cable will run for over 2700 nautical miles from Recife to Las Palmas, and will have some 140 transistor repeaters. Connections between the Canary Islands and the Spanish mainland will be over the PENCAN cables. PENCAN 1 has been in service for 6 years and PENCAN 2, which will carry 1840 circuits, is now being installed. At the Canary Islands BRACAN will also connect with SAT I (360 circuits) to Cape Town and Lisbon. From Lisbon a 640-circuit cable runs to Cornwall, and another 640-circuit cable has recently been opened from Bilbao in Northern Spain to Cornwall. Overland microwave links connect the Iberian Atlantic cable terminals with the Mediterranean cables into Spain (Barcelona-Pisa and Estepona-Rome).

Cable for CANTAT 2 and BRACAN will be made at STC's Southampton plant, the repeaters at North Woolwich and the terminal equipment at Basildon and Newport.

CANTAT 2 is primarily intended for Britain's communications with the North American continent. Since CANTAT 1 was laid the annual total of telephone calls between the U.K. and Canada has increased sevenfold. Calls from Canada to this country now occupy nearly 3½ million minutes a year and from U.K. to Canada about 2 million minutes. Calls from the U.S.A. account for nearly 13½ million minutes a year (compared with less than 2 million in 1960) and to the U.S.A. from Britain over 10½ million minutes (compared with 1½ million in 1960).

Since the construction of CANTAT 1 satellite communications have become established and these are providing an increasing proportion of transatlantic communication every year. But at the same time improvements in the technology of undersea cables allow the number of circuits to be increased greatly without a proportionate increase in cost. Because of this, transatlantic communications between Europe and North America—the busiest inter-continental route in the world—can have new links added economically by submarine cables. For example, whereas CANTAT 1 was laid at a cost of £100 000 a circuit, CANTAT 2 will cost only £16 500 a circuit.

An indication of the future of global telecommunications was given in a recent survey presented by the Post Office at the Telecom 71 exhibition in Geneva recently. At the end of 1970 the U.K. had nearly 2000 telephone-type circuits through submarine cables to other countries. By 1980 these will have



Telephone cables across the North Atlantic Ocean

increased to 11 000 mainly as a result of six additional 1260-circuit cables across the North Sea and additional transatlantic cables, including the second cable to Canada, with 1840 circuits.

Satellite circuits will be up from 290 to 3600 and the capacity of the microwave radio link to France will increase from 1500 to 6400 circuits. As a result, the overall increase of telephone-type circuits between the U.K. and other countries is likely to be from 3750 in 1970 to 21 000 in 1980.

**Cable Design**

The CANTAT 2 cable will be predominantly of lightweight design developed by the Post Office with an outer conductor of aluminium and the strength in a steel rope inside the inner copper conductor. Compared with armoured cables, it is cheaper to produce, easier to lay, and easier to bring to the surface for repair. CANTAT 1 was the first cable of this design. External armour is still used for additional protection on sections of the cable laid in shallow water.

**Repeaters**

The repeaters in the cable have to give a trouble-free life of more than 25 years and each of the CANTAT 2's 490 repeaters (there are 90 in CANTAT 1) is protected by a deep-sea pressure housing of proven design developed by the Post Office and Standard Telephones and Cables Ltd. They use separate amplifiers for each direction of transmission. Distance between repeaters is about 6 nautical miles.

**Transistors**

The transistors in the repeaters were developed by the Post Office and produced by the Post Office and ST&C. A very high standard of reliability is set for them, with a performance standard ensuring that in 25 years' operation less than one transistor in 4000 will fail. About 3200 transistors will be used in the CANTAT 2 repeaters in the main amplifying path. To control the quality of those selected for the cable, another 18 000 transistors will be exhaustively tested electrically and mechanically, some to destruction. The transistors use aluminium wires bonded to aluminium contacts in a process developed by the Post Office Research Department. This bonding has an impressive reliability performance with not a single failure in 40 000 bonds tested in production and 6000 in transistors now on the sea bed.

**Circuit Capacity**

Commercial capacity of 1840 circuits spaced at 3 kHz intervals and arranged as 23 supergroups, use the frequency band 312–6012 kHz in the A–B direction of transmission and 8000–12 700 in the B–A direction. Four service order-wire circuits of nominally 3 kHz bandwidth each use the frequency bands 6024–6036 kHz in the A–B direction and 7976–7798 kHz in the B–A direction of transmission.

**Route**

To select the best route for the cable the *John Cabot* is surveying the Continental shelves on both sides of the Atlantic ocean. Deep ocean areas are being surveyed this summer by a

**Comparisons between the first and second CANTAT cables**

	CANTAT 1	CANTAT 2
Date	1961	1974
Number of speech channels	80	1840
Number of repeaters	90	490
Length in nautical miles	2072	2840
Active elements	480 P.O. type 10P valves	2940 P.O. types 4A and 10A transistors
Principal types of cable	0.99 in. unarmoured (1518 nautical miles) 0.62 in. armoured (554 nautical miles)	1.47 in. unarmoured (2425 nautical miles) 1.47 in. armoured (370 nautical miles)
Power	9.5 kV 415 mA	12.34 kV 500 mA
Cost per circuit between cable terminals	£100 000	£16 500

French cable vessel with British Post Office marine staff on board. In picking the route the Post Office will be helped by detailed information from many sources including the British and Canadian Hydrographic Offices, the National Institute of Oceanography, and the Bedford Institute of Oceanography in Canada. This information is vital to the surveyors picking a path through the valleys and passes of the mid-Atlantic ridge. Other factors that the route planners have to consider include the nature of the sea bed, other cables, and fishing grounds where there is heavy trawling.

**Laying**

The increase in the number of circuits in recent-laid cables has meant closer spacing of repeaters and the Post Office has designed a new cable-laying engine for cable ships. This was first used on the Post Office cable ship *Alert* when the U.K.-Spain cable was laid last year. A modified version of this engine is to be fitted aboard the *John Cabot* and the *Mercury* for the CANTAT 2 task. The *John Cabot* will also be equipped with a plough for burying cable and repeaters in the seabed in shallow areas (less than 300 fathoms) for extra protection.

**Navigation**

The *John Cabot* and the *Mercury* will have to navigate with detailed precision over the main part of the cable route during laying operations. Ships' positions will be fixed by satellite navigation and other radio-navigational systems.

**Correction**

The following amendments should be made to the paper 'Applying the Phase-locked Loop in Communications and Instrumentation', published in the July 1971 issue of *The Radio and Electronic Engineer*:

Page 318, Section 4.4, first line: *delete* the words ' . . . . . from Rugby . . . . . ' .  
third line *should read* 'one part in 10<sup>11</sup>'.