FOUNDED 1925
INCORPORATED BY
ROYAL CHARTER 1961

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

THE RADIO AND ELECTRONIC ENGINEER

The Journal of the Institution of Electronic and Radio Engineers

VOLUME 38 No. 6 DECEMBER 1969

Communicating Facts

"English as a World Language" is the title of the first part of the Annual Report of the British Council for 1968-69. Promotion of the knowledge of the English language abroad is one of the main aims of the Council, effected generally by financing the teaching of English as a second language. Although even the Council is unable to indicate the precise extent to which English is the most used language in the world, it states that English teaching has more hours on the time-table than any other subject. It is the language most used for communicating facts and so a major factor in scientific, technological and economic progress.

Scientists and engineers are well aware that about a half of the world's scientific literature is written in English. The considerable amount of work in all disciplines published in other languages inevitably poses the question: how best to bring the original work in these languages to the notice of those who need to use it? Since scientists and engineers are no more multi-lingual than any other branch of the community, translation is the obvious course and as electronic engineers we think at once of automatic translation using computers. It is fair to say that the early expectations for these techniques have not yet been justified, and although man has solved many technological problems, the answer to the combination of logical operations with illogical language is not just around the corner. Until the etymologists are able to program the idiosyncrasies and nuances of meaning of language we shall continue to find computer translation at the best a rough and ready tool and at the worst the unwitting cause of misunderstanding.

Human translators provide two immediate alternative approaches to the problem: either every paper should be translated into English as the most generally accepted *lingua franca* or translations should be made only when there is a definite request. The 'on-demand' translation, though superficially attractive, runs the real risk of repeating work already done elsewhere; there are schemes for publicizing the existence of a translation but this information is almost impossible to link to the appearance of a paper in the abstract journals and difficult to disseminate fully in other ways later. The first of the alternatives may well be wasteful in employing skilled translators on much material which does not merit consideration; on the other hand 'cover-to-cover' translations of many scientific and technical journals are already produced and the feasibility of extending these arrangements is worth further study.

The efforts of the British Council are obviously of real help to the non-English speaker in learning English but its direct assistance to those speaking only English lies in encouraging the rest of the world to do likewise! The Report points out that dialects of English arise in different parts of the world, which are often incomprehensible to speakers of English as a second language elsewhere, though perhaps not presenting quite such serious problems to those with English as the mother tongue. (One must also recall that Great Britain and the United States of America have been referred to as two countries divided by a common language!) The Council recognizes that, apart from native English teachers, radio, television, and disk and tape recordings can emphasize the international standard of spoken English. The desirability of standardization such as this is a goal familiar to all engineers—the same terms of reference between both sides are essential whatever the nature of the facts to be communicated.

F. W. S.

INSTITUTION NOTICES

Combined Programme of Meetings

The booklet containing details of Institution meetings in London, and meetings of the East Anglian, Kent and Thames Valley Sections for the second half of the current session (February to May 1970) is being sent to all members throughout Great Britain with the January issue of the *Proceedings of the I.E.R.E.* As previously, the booklet presents a combined programme of the above I.E.R.E. meetings together with the London area meetings of the Institution of Electrical Engineers.

Both Institutions wish to re-emphasize that all of their ordinary meetings for which no fee is charged are open to all members of the other Institution. This mutual arrangement extends to members of the Institute of Physics and the Physical Society and of the Institute of Mathematics and its Applications; meetings of these two bodies are also open to I.E.R.E. and I.E.E. members without further formality under the agreement reached by the Standing Committee of Kindred Societies.

Automatic Test Systems Conference Extended

The Organizing Committee for the I.E.R.E. Conference on Automatic Test Systems, to be held at Birmingham University in April 1970 have nearly completed the programme. Over 40 papers will be presented and this has necessitated extending the duration of the Conference by a day beyond the three days originally envisaged. It will now open on Tuesday, 14th April and continue until the afternoon of Friday, 17th April. The outline programme will be published in the Institution's *Journal* and *Proceedings* early in the New Year; requests for further information and registration forms may be made using the application form in the end pages of this issue.

Electronics SDI Service To Start in New Year

The ambitious current awareness project initiated by the National Electronics Research Council and described by Lord Mountbatten in a paper in 1966† is now moving a stage further towards application. It has been announced by INSPEC, the information service in physics, electrotechnology and control operated by the Institution of Electrical Engineers, that it is to launch a selective dissemination of information (SDI) service in electronics in January 1970.

The SDI service will be available on an individual or group subscription basis in the United Kingdom only, and will provide a quick and convenient means of keeping up to date each week with new developments in electronics. Periodical articles on all aspects of electronics, published in English or English translation, will form the basis of the service. It is hoped to start a comprehensive SDI service in 1971 which will cover all languages and the complete subject range of INSPEC.

The I.E.R.E. has maintained its early identification with the SDI project through its association with the I.E.E. and other learned societies in this country and the United States who together make up the INSPEC organization. The new service is part of the overall plan for the development of a comprehensive information service, which is being supported by the Office for Scientific & Technical Information (OSTI) of the Department of Education & Science. For the past year the SDI service in electronics has been limited to some 600 research workers as part of an OSTI-supported information research project.

Further information and details of subscriptions may be obtained from the Manager, INSPEC SDI Investigation, I.E.E., 26 Park Place, Stevenage, Herts.

† Mountbatten of Burma, 'Controlling the information explosion', *The Radio and Electronic Engineer*, **31**, No. 4, pp. 195–208, April 1966.

Institution Giro Account

Members are advised that for the convenience of those who wish to remit their annual subscriptions and other payments through the National Giro, the Institution now has a Giro account. The number is 578 0101 and it may be used in precisely the same manner as any other Giro transaction.

Index to Volume 38

This issue of *The Radio and Electronic Engineer* is the last in the present volume, Volume 38. An Index will be sent out with the February issue.

Members who wish to have their *Journals* bound are asked *not* to send them to the Institution for this purpose before the Index is published.

Reprints of Journal Papers

Reprints are prepared of all papers published in the *Journal* and copies may be obtained from the Institution, price 5s. 0d. each (post free). Requests for reprints may be made using the form which is included in the end pages of most issues of the *Journal*. It is particularly asked that remittances be sent with orders to avoid book-keeping entries and thus reduce handling costs.

U.D.C. 551.55: 534.222.2: 621.396.962.23

An Investigation into the Practicability of Using an Electromagnetic-Acoustic Probe to Detect Air Turbulence

By

D. C. COOPER, Ph.D., C.Eng., M.I.E.E.†

AND

J. BLOGH, M.Sc.†

Presented at a Symposium on 'The Possibility of Detecting Clear Air Turbulence', organized by the Institution's Aerospace, Maritime and Military Systems Group and held in London on 6th March 1969.

Summary: This paper describes some theoretical and experimental investigations into the possible use of the reflexion of electromagnetic waves by an acoustic shock wave. Satisfactory detection of reflected microwave signals has been achieved in a number of experiments and the Doppler frequency shift of the signals has been shown to give a satisfactory indication of the wind velocity along the line of sight of the EMAC (ElectroMagnetic-ACoustic) probe arrangement. Remote indication of radial wind velocity has been achieved at ranges in excess of 450 m (1500 ft) and greater ranges should be obtained with improved experimental equipment.

List of Principal Symbols

A	area of radar receiving aerial
d	spatial distance in the spherical surface
E	electric field
k	wave number $= 2\pi/\lambda$
n	refractive index
P_{trans}	transmitted power
P_{rec}	received power
p	pressure in millibars
p_0	atmospheric pressure
p_p	partial water vapour pressure
p_l	limiting pressure
r	distance from centre of divergence
r_0	distance from sound source to centre of divergence
s	scale of perturbation
T	temperature in degrees Kelvin
T_0	mean temperature
v_s	sound velocity
W_0	$= (\beta/\pi)$. (total reflected power)
Z	characteristic impedance
α	mean square value of phase perturbation of electromagnetic wave
β	$= \left(\frac{ks}{2r}\right)^2 \cdot \frac{1}{\alpha}$
γ	ratio of specific heats
ε_r	relative permittivity
λ	electromagnetic wavelength
λ_a	acoustic wavelength
μ	strength of turbulence
μ_r	relative permeability
ρ	voltage reflexion coefficient
ψ	phase perturbation of acoustic wave

†Department of Electronic and Electrical Engineering, University of Birmingham.

1. Introduction

In the late 1950s some theoretical and experimental investigations were undertaken, by workers in the U.S.A.,[1] to assess the potential value of electromagnetic-acoustic interaction phenomena. It was thought that the phenomena could be used for the measurement of various atmospheric parameters, and in particular, it was hoped that wind speed and turbulence effects could be identified.

The theoretical investigations indicated that the reflexion of electromagnetic waves from acoustic waves in air would provide a rather small reflected signal but the detection of microwave signals reflected from pulses of high-frequency sound was satisfactorily demonstrated with an experimental system arranged as shown in Fig. 1. The sound frequency used in this demonstration was 22 kHz, and the microwave frequency was in the region of 10 GHz, these frequencies being automatically controlled to ensure that the sound wavelength was one-half of the electromagnetic wavelength. In this way the individual signals reflected from the multiple wavefronts of the long pulse of high-frequency sound were made to add constructively and so they provided an adequate overall reflexion.

The first successful demonstration of what has been called the EMAC probe (electromagnetic acoustic) highlighted the difficulties inherent in the use of the technique. Thus extremely high acoustic powers were required, and a microwave radar with a good overall performance was used, but even so the maximum range at which the EMAC interaction could be detected was only about 30 m in calm, or static, air. The range limitation was, in this case, almost entirely due to the rather high attenuation which characterizes the propagation of high-frequency sound waves in air, and this feature together with the difficulty which would undoubtedly be experienced in maintaining the required wavelength ratio when the acoustic wave propagates in turbulent air, has led other workers to consider the application of other forms of acoustic, and electromagnetic modulation.

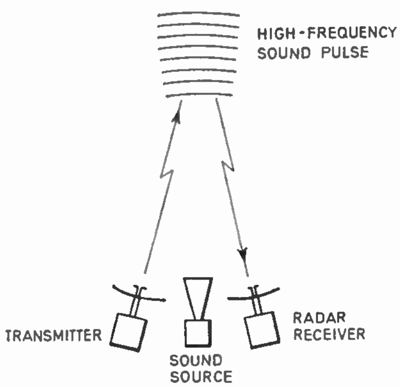


Fig. 1. Demonstration of the reflexion of electromagnetic waves by acoustic waves.

A team in the U.S.A. conducted a comprehensive set of theoretical and experimental studies[2] which were discontinued in about 1965, probably because satisfactory long range operation of the EMAC probe had not been achieved in preliminary trials. These trials utilized shock waves produced by supersonic aircraft and unsuccessful attempts were made to observe the reflexion of electromagnetic waves by using a high-performance pulse radar.

In the summer of 1967 the authors set out to investigate the potentialities of the EMAC probe technique and it was concluded, in the preliminary studies, that the most promising form of acoustic modulation was a shock wave which was capable of propagation without excessive attenuation. A shock wave would also give a satisfactory electromagnetic reflexion coefficient for a reasonable range of electromagnetic wavelengths, or alternatively for a reasonable range of shock wave shapes with a fixed electromagnetic wavelength. At the time of these preliminary investigations we were not aware of the work of the team mentioned above[2] but it is interesting to note that the same basic conclusions were reached in respect of the best form of acoustic wave.

2. The Reflexion of Electromagnetic Waves from a Pressure Discontinuity in the Atmosphere

The partial reflexion of an electromagnetic wave from an acoustic wave in air can be regarded as being caused by the refractive index changes induced by the acoustic wave. In order to deal with the general case of reflexion from an acoustic wave of arbitrary shape it is convenient to start by considering the reflexion coefficient for an electromagnetic plane wave normally incident on the plane interface between two regions of different refractive indices n_1 and n_2. The reflexion coefficient can be obtained by considering that the electromagnetic propagation is from a loss-free region with characteristic impedance,

$$Z_1 = \sqrt{\frac{\varepsilon_0 \varepsilon_{r1}}{\mu_0 \mu_{r1}}} = n_1 \sqrt{\frac{\varepsilon_0}{\mu_0 \mu_{r1}}}$$

towards a loss-free region of characteristic impedance,

$$Z_2 = \sqrt{\frac{\varepsilon_0 \varepsilon_{r2}}{\mu_0 \mu_{r2}}} = n_2 \sqrt{\frac{\varepsilon_0}{\mu_0 \mu_{r2}}}$$

where we have in both cases used the relationship $n = \sqrt{\varepsilon_r}$.

We shall be considering only media for which $\mu_{r1} = \mu_{r2} = 1$ so that the voltage reflexion coefficient is given by

$$\rho = \frac{Z_2 - Z_1}{Z_2 + Z_1} = \frac{n_2 - n_1}{n_2 + n_1} \simeq \frac{\delta n}{2} \qquad \ldots\ldots(1)$$

on the assumption that n_1 and n_2 are very close to unity.

Now the refractive index of air is certainly near to unity and it is given by[3]

$$n = 1 + N.10^{-6} \qquad \ldots\ldots(2)$$

where

$$N = \frac{77{\cdot}6p}{T} + \frac{3{\cdot}73 \times 10^5}{T^2} p_p$$

p = pressure in millibars

p_p = partial water vapour pressure.

and T = temperature in degrees Kelvin

We may in general neglect the second term in the expression for N and on this basis the refractive index can be directly related to the local atmospheric pressure by the expression,

$$n = 1 + \frac{77{\cdot}6p10^{-6}}{T}$$

so that the change in refractive index for a given pressure step can be written as:

$$\delta n = n_2 - n_1 = \frac{77{\cdot}6 \times 10^{-6}}{T} [p_2 - p_1] \quad \ldots\ldots(3)$$

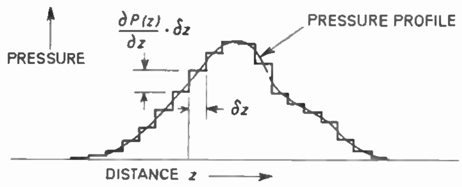


Fig. 2. Step approximation for an arbitrary pressure profile.

If we assume a step change of local pressure of magnitude 1·5 kg/m² (1 lb/ft²), then we estimate the magnitude of the reflexion coefficient by converting the differential pressure into the appropriate units and using equations (1) and (3). In this way we find that,

$$\rho = \frac{\delta n}{2} = \frac{1{\cdot}3\times 10^{-7}}{2} = 6{\cdot}5\times 10^{-8}$$

It may be noted that the above differential pressure of 1·5 kg/m² corresponds to a sound level of 128 dB measured with respect to the standard datum of 0·00002 N/m², and hence it is seen that an extremely small reflexion coefficient is obtained with a sound level which is quite large.

3. Reflexion from Acoustic Waves of Arbitrary Form

In this section we shall consider that the pressure of the acoustic wave taken at any instant of time varies only in the z direction and this direction corresponds to the direction in which the electromagnetic plane wave propagates. Thus the acoustic wave and the electromagnetic waves are both considered plane and are travelling either in the same or in the opposite directions.

To deal with an arbitrary pressure profile for the acoustic wave we assume that this profile may be approximated by a large number of small steps as is illustrated in Fig. 2. Thus we have at the position z the step of pressure given by

$$\delta p = \frac{\partial p(z)}{\partial z}\delta z \qquad \ldots\ldots(4)$$

where $p(z)$ is the pressure expressed as a function of z.

Now the above pressure discontinuity will give rise to a small contribution to the overall reflexion coefficient, the contribution being given by

$$\delta\rho = L\delta p$$

where we see from equations (1) and (3) that

$$L = \frac{77{\cdot}6}{2T}\times 10^{-6}$$

Providing that multiple reflexions can be neglected, the overall reflexion coefficient for the step approximation to the pressure profile can be obtained as the sum of the individual contributions with phasing terms introduced to take account of the position along the z direction at which each contribution originates. Since in our case the individual contributions will be extremely small, multiple reflexion effects will certainly be negligible and hence the overall reflexion coefficient will be

$$\rho = \sum L\delta p\, \mathrm{e}^{-2jkz}$$
$$= \sum L \frac{\partial p(z)}{\partial z}\,\delta z\, \mathrm{e}^{-2jkz}$$

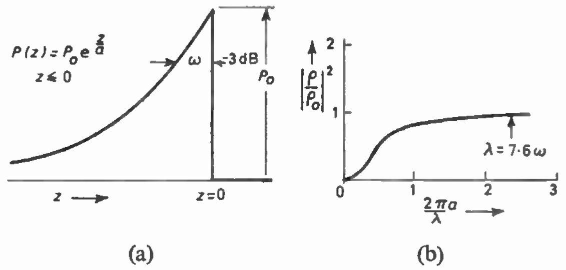
where $k = 2\pi/\lambda$, and λ = wavelength of the electromagnetic waves.

In the limit the above sum approaches the integral,

$$\rho = L\int \frac{\partial p(z)}{\partial z}\,\mathrm{e}^{-2jkz}\,\mathrm{d}z \qquad \ldots\ldots(5)$$

which is the desired expression for the voltage reflexion coefficient.

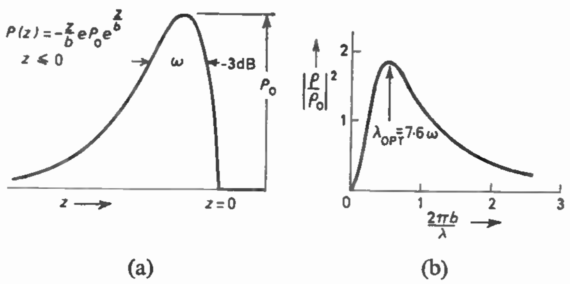
In order to illustrate how the sound pressure profile affects the magnitude of the electromagnetic reflexion coefficient we show in Figs. 3, 4 and 5 the manner in which the power reflexion coefficient varies with wavelength for three pressure profiles. These profiles are reasonable approximations to those recorded experimentally near to an explosive source, at a moderate distance from the source, and at a considerable distance.



(a) Pressure profile.

(b) Plot of normalized power reflexion coefficient against $2\pi a/\lambda$.

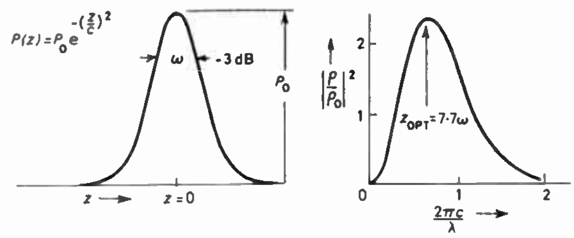
Fig. 3. Reflexion at short range.



(a) Pressure profile.

(b) Plot of normalized power reflexion coefficient against $2\pi b/\lambda$.

Fig. 4. Reflexion at medium range.



(a) Pressure profile.
(b) Plot of normalized power reflexion coefficient against $2\pi c/\lambda$.
Fig. 5. Reflexion at long range.

The power reflexion coefficients shown in Figs. 3(b), 4(b) and 5(b), have been normalized with respect to the power reflexion coefficient for a pressure step of magnitude equal to the peak excursion of the pressure profiles, and it will be observed that for optimum operating conditions the normalized power reflexion coefficient does not differ very much from unity. Furthermore, it is apparent that an optimum value for the electromagnetic wavelength is about 7·7 times the -3 dB width of the acoustic pressure profile.

It is well known that a single shock wave is not a stable profile and after propagation over a considerable distance a stable 'N' wave profile is generated. The reflexion coefficient for a profile of the 'N' type can be obtained by considering it to be composed of two single shock waves spaced by an appropriate distance, and it is obvious that the overall reflexion coefficient may be greater or less than that for a single shock wave. We have not conducted any experiments at ranges where the 'N' wave has formed and hence we use only the results obtained for single shock waves.

4. The Reflexion of Electromagnetic Waves by an Almost Spherical Surface

We now consider the problem of calculating the signal received when a radar is placed at or near a shock wave source and the reflecting region can be regarded as being nearly spherical. For convenience we assume that the radar aerial can be represented by an aperture in which a given electromagnetic field is excited, but this does not restrict the ideas and results we obtain to this particular form of electromagnetic radiating structure.

The general situation is depicted in Fig. 6 and the first stage in the analysis is the calculation of the incident field strength of the electromagnetic waves at a perfectly spherical surface. For the purpose of this calculation we assume that the aerial aperture is excited so that the electric field has only a component $E_y(x_1, y_1)$ in the y_1 direction at all points in the aperture, and we also assume that the aerial aperture is large compared with the wavelength of the electromagnetic radiation so that the far-field radiation pattern is restricted to a narrow range of angles close to the normal to the aperture.

The initial calculations will assume a spherical reflecting surface which is centred on the point $x_1 = 0$, $y_1 = 0$. We use the rectangular co-ordinate system x_2, y_2 to define positions on the spherical surface which can be regarded as being very nearly a plane surface in the region where it is illuminated by the narrow far-field radiation pattern of the aerial. However, despite this the distance of any point on the x_2, y_2 surface from the origin of plane x_1, y_1 is precisely the radius of the spherical surface. From these assumptions it is clear that we are limiting our attention to values of r which place the reflecting surface in the far-field region of the radar aerial.

The incident field at plane 2 will have only a y_2 component of electric field and this can be evaluated by integrating the contributions produced at each point (x_2, y_2) by all the elements of the aerial aperture, with the appropriate phase allowance being made for the various distances from the aperture elements to the point (x_2, y_2). Thus we obtain

$$Ey(x_2, y_2) = \int\int \frac{\mathrm{j}\exp\left\{-\mathrm{j}k\left(r - \frac{x_1' x_2 + y_1' y_2}{r}\right)\right\}}{\lambda r} \times Ey(x_1', y_1')\,\mathrm{d}x_1'\,\mathrm{d}y_1' \qquad \ldots\ldots(6)$$

where x_1' and y_1' are co-ordinates of position in plane 1 and the integration covers the entire aerial aperture.

Assuming that the field reflexion coefficient at all points of the spherical surface is ρ we now compute the received field in plane 1 by a similar process of integrating all the reflected field contributions. We note that the assumption of a constant reflexion coefficient implies that the radiation of the shock wave is over a very much larger range of angles than the radiation beamwidths of the radar aerial and this is the normal practical situation.

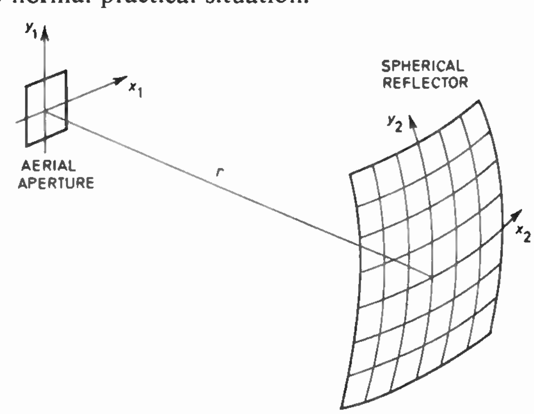


Fig. 6. Reflexion from a spherical surface.

The received field is therefore given by:

$$Ey_{rec}(x_1, y_1) = \iint \frac{\mathrm{j}\exp\left\{-\mathrm{j}k\left(r - \frac{x_1x_2+y_1y_2}{r}\right)\right\}\rho E_2(x_2, y_2)\,\mathrm{d}x_2\,\mathrm{d}y_2}{\lambda r}$$

$$= -\frac{\rho\,\mathrm{e}^{-2\mathrm{j}kr}}{k^2\lambda^2}\iint\left[\iint Ey(x_1', y_1')\exp\left\{\mathrm{j}k\left(\frac{x_1'x_2+y_1'y_2}{r}\right)\right\}\mathrm{d}x_1'\,\mathrm{d}y_1'\right]$$

$$\times\exp\left\{k\left(\frac{x_1x_2+y_1y_2}{r}\right)\right\}\mathrm{d}\frac{kx_2}{r}\,\mathrm{d}\frac{ky_2}{r} \qquad \ldots\ldots(7)$$

Now the Fourier transform relationship for functions of two variables can be expressed by the identity,

$$F(x, y) = \frac{1}{4\pi^2}\iint_{-\infty}^{\infty}\left[\iint F(x'y')\exp\{\mathrm{j}(\omega_1x'+\omega_2y')\}\,\mathrm{d}x'\,\mathrm{d}y'\right]\exp\{-\mathrm{j}(\omega_1x+\omega_2y)\}\,\mathrm{d}\omega_1\,\mathrm{d}\omega_2 \qquad \ldots\ldots(8)$$

and by comparing this identity with equation (7), and noting that $k^2\lambda^2$ is equal to $4\pi^2$, we easily see that,

$$Ey_{rec}(x_1, y_1) = -\rho\,\mathrm{e}^{-2\mathrm{j}kr}Ey(-x_1, -y_1) \qquad \ldots\ldots(9)$$

Thus, we see that under the conditions assumed in the above analysis the received field at any point is a constant factor times the radiated field at the diametrically opposite position in the plane. Provided that the aerial aperture has diametrical symmetry, and all the reflected energy which falls on the aperture is absorbed by the radar receiver, the overall power loss between transmission and reception is given by the factor $|\rho|^2$. This loss is not directly affected by the range to the reflecting surface although in the EMAC probe application the value of $|\rho|^2$ will usually decrease as range increases owing to the propagation losses of the shock wave.

It is impracticable to place a single aerial at the same position as the source of a shock wave of high intensity and we see from equation (9) that a satisfactory arrangement is to put a transmitting aerial and a receiving aerial at points equally-spaced on opposite sides of the sound source. This arrangement is also consistent with the idea of specular reflexion from the spherical surface of the shock wave.

We now turn our attention to the case in which perturbations appear in the surface. The simplest way to deal with this problem is, we believe, to utilize the fact that in dealing with the ideal case we have found that the received field at the aerial plane is proportional to the Fourier transform of the reflected field at the spherical surface. This, together with the fact that the field incident on the spherical surface is proportional to the Fourier transform of the radiated field in the aerial plane, is the reason for the simple inversion of field pattern which appears, together with the reflexion coefficient, in the expression (eqn. (9)) relating the radiated and received fields.

Now if the spherical surface is perturbed in position in a given manner we can allow for this by introducing a phase factor into the reflected field expression before performing the second Fourier transformation to obtain the received field distribution. However, the same result can be obtained in an alternative fashion because it is well known that the Fourier transform for the product of two functions is given by the convolution of the Fourier transforms of the individual functions. The Fourier transform of one of our functions is obviously the received field for the ideal spherical reflector case, and all we have to do is to obtain the Fourier transform for the phase perturbation function which represents the physical perturbation of the reflecting surface, and then carry out the convolution process to obtain the received field under the perturbed conditions.

The simplest form of perturbation is a tilt of the reflecting surface giving a phase perturbation function which is just a linear phase slope in the appropriate direction. The Fourier transform of such a phase function is a delta function offset from the origin, and since the convolution of any function with a delta function just translates the original function by an amount corresponding to the position of the delta function, the received field is that which would have been received in the ideal case but it is translated to a new position. It is not difficult to show that this received field position satisfies our intuitive ideas concerning specular reflexion and the situation is illustrated in Fig. 7.

When turbulence is present in the atmosphere the acoustic shock wave will have random perturbations from the ideal spherical form and we will assume that any fluctuations in the intensity of the shock wave are negligibly small. Thus we interest ourselves in the effect of the random phase perturbations on the received signal.

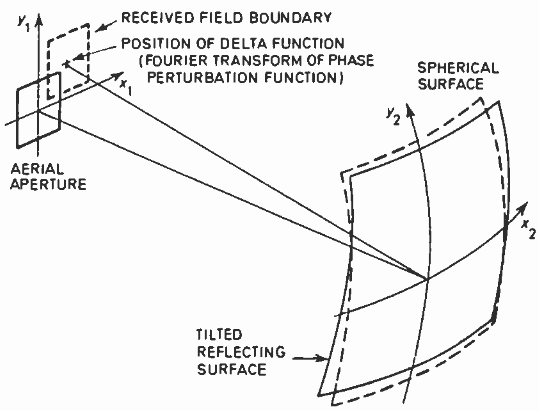


Fig. 7. Reflexion from a tilted spherical surface.

The problem is similar to the investigation of the spread of intensity in an optical image which has been analysed by Chernov.[4] Chernov assumes that there are both amplitude and phase fluctuations in the illumination of the converging spherical wavefront which is produced by a converging lens. With no perturbations the amplitude and phase are constant on the spherical surface and the lens produces an image which is very small, almost a point.

The analysis given by Chernov assumes that the fluctuations have Gaussian probability density functions and spatial autocorrelation functions of normalized form given by e^{-d^2/s^2}, where d is the spatial distance in the spherical surface and s is a measure of the scale of the perturbations. Applying Chernov's result to our problem we find that the power density distribution in the (x_1, y_1) plane, this corresponding to the focal plane in the optical case, is given by:

$$W = W_0\, e^{-\beta(x_1^2+y_1^2)} \qquad \ldots\ldots(10)$$

where $W_0 = (\beta/\pi)$. (total reflected power)

$$\beta = \left(\frac{ks}{2r}\right)^2 \frac{1}{\alpha}$$

and α = mean square value of phase perturbation.

It should be emphasized that the above distribution actually describes the defocusing effect produced by the random perturbations. However, for the case in which the ideally-focused field pattern is of small size and the distribution given by equation (10) is of considerably greater size, it is sufficient for us to regard equation (10) as giving an accurate indication of the actual power density distribution.

We can now easily estimate the total received power in the presence of turbulence which produces a given value of mean square phase perturbation in the reflected electromagnetic waves. Thus, if it is assumed that the shock wavefront is not tilted and that the transmitting and receiving aerials are correctly sited, the total received power will be $W_0 \times$ (area of receiving aerial).

If a tilt of the shock wave has occurred or if incorrect aerial positioning has been used, the total received power will be given by

$$P_{rec} = W_0\,(\text{receiving area}).\, e^{-\beta r_1^2} \qquad \ldots\ldots(11)$$

where r_1 is the distance from the receiving aerial to the position of maximum power density.

The point of maximum power density can be found by using geometrical optics, that is, by assuming that the reflexion is specular from the unperturbed reflecting surface.

It will be noted that we have not considered the possibility of the radar aerials being incorrectly positioned along the line of sight or axis of the EMAC probe system. Chernov[4] has also considered the light intensity distribution along the axis of an optical system and if we use his results for the EMAC probe configuration in which a narrow beam radar system is used we find that an axial positioning error will produce considerably less loss of received power than a similar error in the transverse direction. It is, therefore, reasonable to neglect the effects of the axial positioning errors.

5. The Propagation of Acoustic Waves in Air

An acoustic wave propagating through the atmosphere suffers attenuation owing to spherical divergence, absorption and scattering along its path. Absorption is produced by viscous losses, heat radiation and conduction, and the diffusion of molecules from the fast moving regions of the wave into the slowly moving regions. Another factor is molecular absorption which occurs when water molecules are set in motion by the acoustic wave and absorb energy from it.

Inhomogeneities in the atmosphere are the cause of scattering but in practice the attenuation so produced is negligible, being of the order of 0·003 dB/m (0·001 dB/ft).

The total absorption of an acoustic wave is very much influenced by the frequency. At a frequency of 2 kHz the attenuation factor is of the order of 0·015 dB/m to 0·03 dB/m depending on the relative humidity, and for a frequency of 10 kHz the attenuation factor will be in the range of 0·2 dB/m to 0·4 dB/m. It is obvious that propagation with small loss can only be achieved at low frequencies and the rather high attenuation factor corresponding to a frequency of about 22 kHz is the main reason for the limited range of operation of the Mid West system.[1]

A large amplitude acoustic wave of arbitrary profile will progressively change its shape as it propagates, so that it ultimately has a triangular shock wave shape, or alternatively becomes a series of triangular shock fronts. This change of profile is induced by the differential velocities of various regions of the wave. High-pressure regions travel with a higher velocity than do low-pressure regions owing to two phenomena, first the differential particle velocity between high-pressure and low-pressure regions and second the increased velocity of sound in the high-pressure regions caused by the local adiabatically increased temperature.

The above factors cause the wavefront to become progressively steeper but this process cannot lead to a vertical wavefront because of compensating effects in the form of thermal conduction across the shock region and viscous forces produced by the rapid change of pressure.

These compensating effects limit the shock wave slope and amplitude by transferring heat and momentum from the high-pressure, high-temperature regions of the wave to the other regions, and in this way a balance is established between the steepening action of the different propagation velocities and the compensating dissipative effects.

Under the above conditions it can be shown[5] that with an arbitrarily large sound level produced by a source, the sound level obtained at a distance from the source will have a finite amplitude limit which cannot be exceeded by any increase in the source level.

For a plane wave the limiting pressure is given by the expression

$$p_L = \frac{\gamma \lambda_a p_0}{(\gamma+1)r} \qquad \ldots\ldots(12)$$

where p_0 = atmospheric pressure
λ_a = acoustic wavelength
r = distance from source

and for a spherical wave,

$$p_L = \frac{\gamma \lambda_a p_0}{(\gamma+1)r \log_e \dfrac{r}{r_0}} \qquad \ldots\ldots(13)$$

where r = distance from centre of divergence
r_0 = distance from sound source to centre of divergence.

It is fairly obvious that the limiting pressure falls very rapidly as the distance from the source is increased and this finite amplitude effect would appear to represent a fundamental limitation on the range which can be obtained in an EMAC probe system. The limitation certainly applies if the source itself produces high level shock waves, and this is the case for the experiments described later in this paper. However, it has been suggested that high-amplitude shock waves can be produced at a large range from a source which is designed to generate an acoustic sound pulse containing only low-frequency components. In this case one generates a long pulse, say of Gaussian shape, which has a large energy content but only a modest amplitude, and the differential velocity effects occurring during propagation ultimately result in the formation of a large amplitude shock wave. This approach seems to be an efficient method for producing shock waves only where they are required.

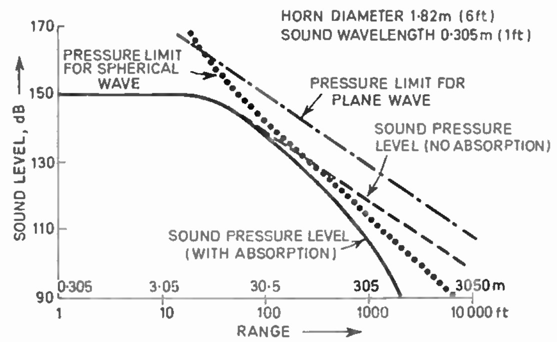


Fig. 8. Finite amplitude sound limits.

Returning to the high-amplitude source method for shock wave generation, we indicate in Fig. 8 the limiting pressures which are given by equations (12) and (13) assuming source parameters appropriate to the experimental sources used in our experiments.

Sound pressure levels for the shock waves produced by two explosive sources, a 0·22 in calibre starting pistol and a 10-bore yachting cannon fitted with a 2 m diameter exponential horn, are shown in Fig. 9. It will be observed that the cannon does appear to produce a shock wave which is close to the finite amplitude limit at distances above 30 m, but the shock wave produced by the 0·22 in calibre pistol

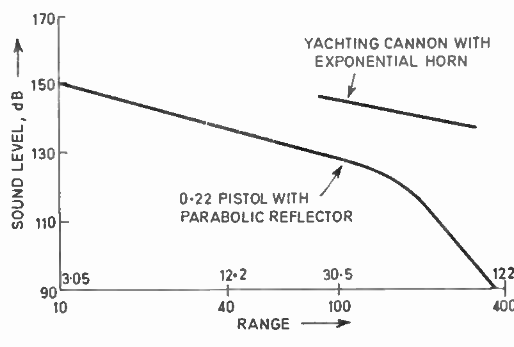
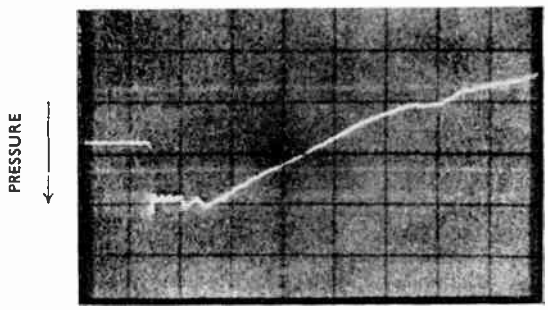
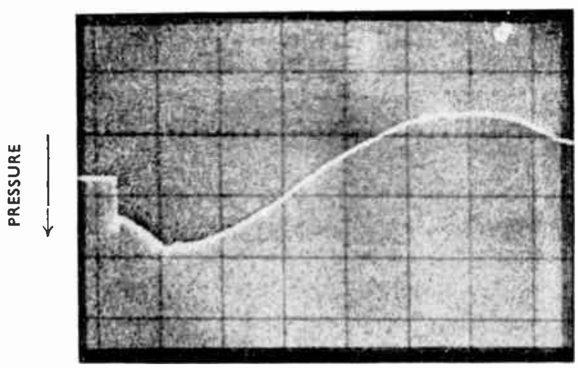


Fig. 9. Measured sound levels.

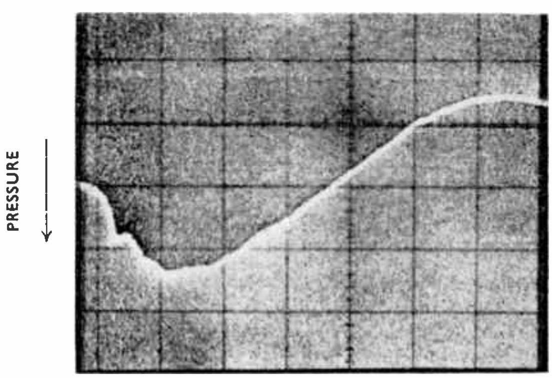
falls well below the finite amplitude limit. The cannon does seem to be an adequate source for ranges up to about 30 m in view of the fact that no shock wave improvement can be expected from an increase in source sound level. Typical pressure profiles for the cannon are shown in Fig. 10.



(a)



(b)



(c)

(a) Range 100 ft
(b) Range 200 ft
(c) Range 400 ft
Time scales 0·5 ms/division

Fig. 10. Shock waves produced by cannon with horn.

Before we consider the effect of atmospheric turbulence on the propagation of the acoustic wave it is important that we recognize that steady winds can have a marked effect on the performance of the EMAC probe. The most troublesome steady wind is one which is perpendicular to the line of sight of the system because in this case the spherical shock front is carried downstream by the wind and the optimum position for the radar aerials may be at some considerable distance from the sound source. This distance will be different for various ranges of the shock wave, because the downstream drift is greater for propagation to the longer ranges, and it does not seem feasible to set up a system which is optimum for a wide variety of cross wind speeds and ranges. The cross wind situation is illustrated in Fig. 11.

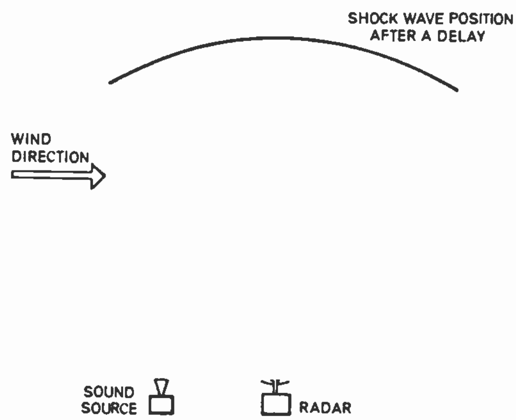


Fig. 11. Effect of a cross wind.

Steady winds along the line of sight produce some defocusing of the received signals in the source plane owing to the fact that the centre of the shock wave front will be either in front of or behind the source. This effect can also be allowed for in the siting of the radar aerials but an arrangement which achieves optimum performance for a range of conditions is again not really feasible.

Non-uniform winds with various velocity gradients have been considered by Allen and Weiner[5] and they obtain results which give the shape of the acoustic shock wave front at various ranges. The results do of course depend on the velocity profiles considered but in general the resulting wavefront is very nearly spherical and will not normally be centred on the source. We again note that optimum siting of the source and the radar aerials is not practically feasible for a wide range of atmospheric conditions.

In view of the fact that perturbations in the shock wave surface produce a spreading of the reflected electromagnetic field about the optimum receiving aerial position we expect to obtain acceptable signals for non-optimum aerial positions providing that atmospheric turbulence exists, and we now consider the magnitude of the surface perturbations which appear under turbulent conditions.

The propagation of waves in turbulent media is the main theme of the work of Chernov[4] and he has demonstrated that the mean square value of the phase perturbations of an acoustic wave of wavelength λ_a which propagates a distance r in a turbulent medium is given by

$$\langle\psi^2\rangle = \frac{2\pi^2\langle\mu^2\rangle sr}{\lambda_a^2} \qquad \ldots\ldots(14)$$

where

$$\mu = \frac{\Delta V}{v_s} + \frac{1}{2}\frac{\Delta T}{T_0} = \text{strength of the turbulence}$$

ΔV = wind speed variation

v_s = velocity of sound

ΔT = temperature variation

T_0 = mean temperature

and s = scale of the turbulence.

In deriving the above result Chernov assumes that μ has a Gaussian probability density function with spatial auto-correlation of normalized form e^{-d^2/s^2} where d is spatial distance, and he shows that the scale of the perturbations corresponds to the scale of the turbulence s.

The conditions assumed for the derivation of eqn. (14) are consistent with those used in dealing with the de-focusing effects of shock wavefront perturbations, and therefore eqns. (10), (11) and (14) can be used for the estimation of performance for an EMAC probe assuming specific atmospheric conditions. The conditions which must be specified are the mean wind speed and its direction, the strength of turbulence as determined by the parameter μ, and the scale of the turbulence determined by the parameter s.

6. Estimation of the Performance of an EMAC Probe System

We are now in a position to estimate the range at which useful electromagnetic reflexions will be received from an acoustic shock wave. The sound source for the system will be taken to be the 10-bore yachting cannon which provides a shock wave whose intensity as a function of range has been given in Fig. 9.

Under very calm atmospheric conditions and with correct radar aerial positioning the received power can readily be obtained for any range simply by evaluating the power reflexion coefficient appropriate to the shock wave intensity at that range. However, the above conditions give a maximum signal level and we are particularly concerned with the reduction in signal strength which occurs when air turbulence exists in the path of the acoustic wave.

If we use eqn. (14) to obtain an expression for the mean square phase perturbation of the reflected electromagnetic wave and then substitute in eqn. (11), we get an expression for the received power in the presence of turbulence and cross wind which is as follows:

$$P_{rec} = P_{trans}|\rho^2|A.\ \frac{s}{4\pi\sqrt{\pi\,\overline{\mu^2}}\,R^3}\ \exp\left\{-\frac{s(x_1^2+y_1^2)}{4\sqrt{\pi\,\overline{\mu^2}}\,R^3}\right\} \qquad \ldots\ldots(15)$$

where A is the area of radar receiving aerial.

In the absence of turbulence and cross winds the received power will of course be $|\rho^2|P_{trans}$, assuming that optimum radar aerial positioning is used, and thus the factor

$$\frac{A\,s}{4\pi\sqrt{\pi\,\overline{\mu^2}}\,R^3}\ \exp\left\{-\frac{s(x_1^2+y_1^2)}{4\sqrt{\pi\,\overline{\mu^2}}\,R^3}\right\} \qquad \ldots\ldots(16)$$

gives the loss of received power caused by the atmospheric conditions. Thus all we need to define for the estimation of the turbulence and cross wind effects are the scale of the turbulence s, the turbulence strength μ and the distance r_1 (i.e. $\sqrt{(x_1^2+y_1^2)}$), of the radar receiving aerial from the optimum position. If the incorrect positioning is entirely the result of cross wind effects it is easily found that

$$r_1 = 2R\left[\frac{\text{cross wind velocity}}{\text{sound velocity}}\right]$$

If the cross wind velocity is taken to be zero and we select fairly typical values for $\mu_{r.m.s.}$ and s, say $\mu_{r.m.s.} = 0{\cdot}01$ and $s = 30$ m (100 ft), and if we also assume that the radar receiving aerial has an area of $0{\cdot}93\ m^2$ (10 ft^2) we find that the loss in received power will be approximately 25 dB for a shock wave at a range of 152 m (500 ft) and about 43 dB for a range of 610 m (2000 ft).

The overall loss for the EMAC probe system using the 10-bore yachting cannon is illustrated in Fig. 12 in which the loss in calm air with optimum system geometry is estimated from the measured shock wave data and the additional loss due to turbulence is calculated for the conditions mentioned above.

7. Experimental Results

The radar used in the experimental trials was a 3 cm continuous wave Doppler radar equipped with a Doppler filter bank which provided velocity information to an accuracy better than 2 knots.

The radar was normally operated in a bistatic mode, the transmitting and receiving aerials for this arrangement being placed approximately 3 m on opposite sides of the cannon and its exponential horn. The entire system was directed at various elevation angles between 5° and 50° during a series of tests.

Since the radar operating wavelength was 3 cm the experiments were not conducted under conditions which would optimize the reflexion coefficient for the shock waves whose profiles have been illustrated in Fig. 10. A considerably longer operating wavelength would be required for optimum reflexion and because of this it was anticipated that reflexions would not be observed at ranges much above 300 m (1000 ft) even under calm conditions. This conclusion, which allows for a radar overall performance of 180 dB, can easily be checked by referring to Fig. 12.

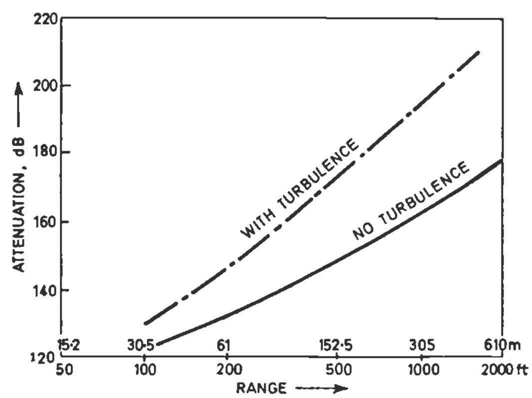


Fig. 12. Graph of attenuation of received signal against range with cannon and horn as sound source.

The results of the experimental trials were basically in agreement with the above conclusion. Thus although during several trials the reflexions were observed at ranges up to about 600 m (2000 ft), the majority of the trials gave reflexions which could only be detected at ranges up to about 300 m.

A limited number of early trials were conducted with the radar using a single aerial for both transmission and reception, this aerial being situated about 3 m from the sound source. The trials made using this monostatic arrangement were considerably less successful than those using the bistatic arrangement and this indicates that the use of correct system geometry is an important factor in obtaining satisfactory performance.

For observation purposes the outputs of the radar filter bank were sampled sequentially at a high sampling rate and the sampled data were displayed in intensity-modulated form on a cathode-ray oscilloscope which had a long term storage facility. The use of a stored display was most valuable because the signals obtained during each trial only lasted for about one second and an accurate visual assessment of the signal frequencies could not be made in this rather short time.

The form of the display is shown by the photograph of Fig. 13. As mentioned previously the sampled data are presented in intensity form and a rapid vertical deflexion of the oscilloscope trace is synchronized to the sampling switch so that the output of any one filter is always presented at a fixed position on the vertical scale. The trace is also given a horizontal sweep of constant velocity, this sweep being triggered off by the sound source, and hence the horizontal scale indicates the time after the initiation of the shock wave. However this time scale can readily be interpreted as the range of the shock wave with a reasonable degree of accuracy and this has been done in the labelling of the scale shown in Fig. 13.

The presence of a detectable signal is indicated by an almost horizontal line such as appears in Fig. 13. Obviously the trace should fluctuate above and below a horizontal line when turbulence is present since this will speed up or slow down the shock wave at various points along its path. This effect cannot be observed in Fig. 13 because the vertical scale corresponds in length to a velocity range of 650 km/h (350 knots) and obviously only very large wind speed variations will produce a noticeable effect.

Expansion of the vertical scale of the display will of course make the effects of moderate wind speed fluctuations apparent to the observer, but for the purpose of the trials an alternative method was used for obtaining the radial wind velocity data. This method relied on the use of an instrumentation tape recorder to record the signals received by the radar. It was then possible to analyse the recordings using standard techniques to obtain very precise Doppler frequency data. A typical curve of radial wind speed obtained in this way is shown in Fig. 14.

As a result of a considerable number of trials it is concluded that the EMAC probe technique does provide reasonably accurate information about the radial wind velocity at remote points, which in the experiments were up to 300 m distant. Occasionally data could be obtained at ranges up to 600 m and again there is no doubt that these data were reasonably accurate. The trials were conducted in a variety of conditions ranging from a virtually static atmospheric to one in which gusts of up to 70 km/h (37 knots) were recorded.

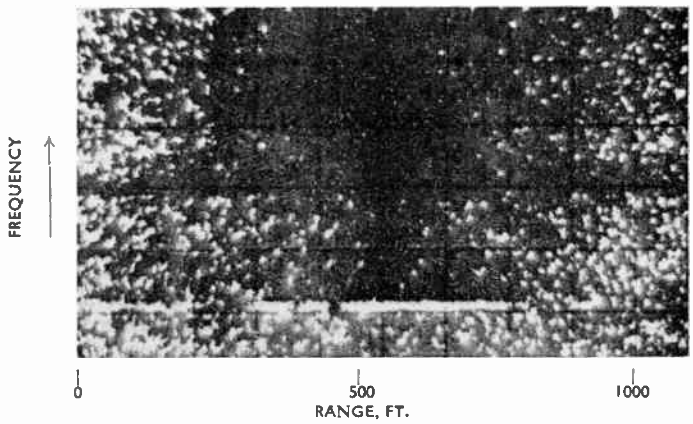


Fig. 13. Display of frequency of signal reflected by acoustic shock wave.

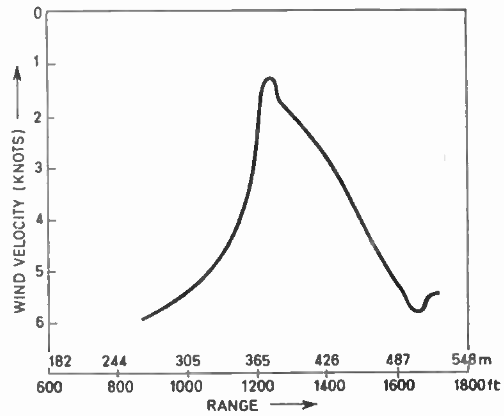


Fig. 14. Radial wind velocity versus range.

8. Conclusions

In this paper we have shown that the EMAC probe technique is useful for remote wind velocity sensing and the range performance of an experimental system has been found to be consistent with our theoretical predictions. The range performance that has so far been achieved is basically of the order of 300 m but there is no doubt that this can be increased with the help of better radar performance and the use of the optimum radar frequency, together with the application of a more efficient shock wave source.

9. Acknowledgments

The authors wish to record their appreciation of the generous assistance which was provided by members of the Royal Radar Establishment, particularly C. Gredley and D. Cheale.

10. References

1. Fetter, R. W., Smith, P. L. and Klein, V., 'Design Concept Study for a Wind Profile Instrument', Mid West Research Institute Project No. 2413-E. Final Report, November 1960.
2. Allen, C. H. and Weiner, S. D., 'Preliminary Experimental Study of EMAC Probe Using Acoustic Shock Waves as Reflecting Surfaces', Bolt, Beranek and Newman Inc., Report No. 1228, November 1965.
3. Johnson, J. C., 'Physical Meteorology'. (Technical Press of M.I.T., Cambridge, Mass., 1960).
4. Chernov, L. A., 'Wave Propagation in a Random Medium', pp. 125–146, (McGraw-Hill, New York, 1960).
5. Allen, C. H. and Weiner, S. D., 'Study Directed Toward Optimization of Operating Parameters of the EMAC Probe for the Remote Measurement of Atmospheric Parameters', Bolt, Beranek and Newman Inc., Project 66 72, Task 66705, Final Report, September 1963.

Manuscript first received by the Institution on 16th June 1969 and in final form on 6th October 1969. (Paper No. 1290/AMMS25)

The Authors

Dr. Dennis C. Cooper has spent much of his academic and professional life at the University of Birmingham. He received his B.Sc. degree in electrical engineering in 1947, returning eleven years later to take an M.Sc. degree in information engineering in 1959 and he gained his doctorate in 1961; during this second period at Birmingham he held a Whitworth Fellowship. His industrial experience was obtained with Associated Electrical Industries Ltd. at Manchester where he worked on microwave engineering projects between 1947 and 1958. Dr. Cooper is at present a senior lecturer in the Electronic and Electrical Engineering Department of the University of Birmingham. He has previously contributed several papers on radar and associated topics to the Institution's *Journal.*

Julian Blogh obtained a B.A.I. degree from Trinity College, Dublin, in 1965 and an M.Sc. in information and systems engineering in 1967 from the University of Birmingham. During the period 1965–66 he worked as a research engineer in the Radar Research Laboratory of Elliott Bros. at Borehamwood. He is currently an engineer with Ferranti Ltd., Edinburgh, and is engaged in research work within the Electronic and Electrical Engineering Department of the University of Birmingham.

LETTERS

Paper Transistors

Permit me to comment on the letter by Dr. Eisler, on the subject of 'Paper Transistors', which appeared in your March 1969 issue.†

Dr. Eisler proposes a 'subtractive' instead of an 'additive' method for forming thin film integrated circuits, and claims in summary that: 1. The 'additive' approach is only suitable for the fabrication of vast quantities of identical circuits (and the reason Westinghouse is developing the process is 'to consolidate the monopoly of the market by the giant firms'). 2. 'Methodologically' our approach is an error, and somehow betrays the printing tradition. 3. Only the subtractive method is capable of producing the 'millions of different circuits' needed in small quantities. 4. The multi-layer thin film laminate proposed by Dr. Eisler would form a suitable 'raw material' for the subtractive fabrication, e.g. by etching, of complete integrated circuits.

With due respect to Dr. Eisler's contributions to printed circuit technology, I have to characterize his arguments as naive in the extreme, and none of them will stand up to serious scrutiny. The subject is too large, of course, to allow an adequate discussion here, but let me try to cover some of the main issues.

I will start with point No. 2, as this can be disposed of briefly. Whatever the printing tradition may be, printing is clearly not a subtractive method. On the contrary, from the earliest days, all the printing techniques have in fact been 'additive', and are so at the present time. Letterpress, lithography, gravure—all add material (namely ink), in selected locations to the substrate. No printing technique known to me uses a subtractive procedure in forming the final product and such a hypothetical procedure would be quite hard to visualize, let alone put into practice. Of course, subtraction is used in the preparation, e.g., of photogravure printing plates or halftone letterpress blocks, and likewise, we use subtractive methods in preparing our deposition masks.

As for 'methodological' errors, I feel these cannot be discussed in the abstract. If Dr. Eisler can propose (or better still, demonstrate) a superior subtractive approach to our goal, I shall be among the first to embrace his techniques. However, he completely fails to do this in his letter, or, indeed, in his book, and this brings me to the major points of contention, claims No. 3 and 4 above. I will leave No. 1 until last, as it touches on political as well as technical issues.

Integrated thin film circuits, to be generally useful, must contain passive elements (resistive, capacitive and preferably also inductive), active elements (diodes, amplifiers, bistable elements) interconnexions and terminations suitable for interfacing with the outside world. To imagine that the very diverse, and mostly contradictory material and thickness requirements of these components could be simultaneously satisfied by a single multi-layer laminate (or even a range of such laminates) can only be described as very unrealistic wishful thinking. Furthermore, the very real difficulties of high resolution geometric delineation remain the same as with the additive process, and many desirable geometric configurations (e.g. a larger area of material A covering and overlapping a smaller area of material B) cannot be achieved at all. I consider Dr. Eisler's 'universal' laminates an unattainable and unfruitful concept. One of the major advantages of a thin-film process is the availability of a tremendous range of materials, and one's consequent ability to select specific materials or often, combinations of materials, for specific device purposes. It is precisely by virtue of this unlimited repertoire of materials and phenomena, combined with the similarly unlimited degree of geometrical freedom inherent in a synthesizing process that thin-film electronics has such attractive potential. Dr. Eisler's putative subtractive process would negate both these major advantages, without offering any compensating features in return. So far from liberating the art of electronics from the tyranny of standardized circuits, he would impose the far more restrictive tyranny of his multi-layer laminates. If such laminates become technically feasible—which I consider exceedingly unlikely—they will have to be very complex structures indeed. I shudder to think of the highly monopolistic position which would be acquired by the 'giant firms' which would be the only ones in the possession of the complex technology needed to develop and fabricate such 'raw material'. One has to face the fact that electronic circuits, even the simplest ones, are quite complex structures, and this irreducible complexity has to enter at some point in the fabrication process—it cannot be wished away.

Since Dr. Eisler fails to show how even the simplest circuit, or, indeed, a single transistor, could be produced by selective etching of a multi-layer laminate, I cannot consider seriously his claim that such a process would be suitable for producing millions of different circuits.

Let me however agree with Dr. Eisler on one point. The pressure towards circuit standardization, although not without merit, arises far more from the IC-manufacturers' desire (and need) for long runs, than from the users' needs, who will continue to crave variety. This is not a new situation in the history of technology. I predict that the conflict will intensify, and the pressure will increase for the development of alternative technologies better able to supply the diversity of circuits needed, and do so at reasonable cost. This pressure is already producing a variety of 'hybrid' solutions, and is one of the main motivations for our work.

This brings me to the discussion of statement No. 1: '. . . suitable only for the production of vast quantities of identical circuits . . . giant firms . . . monopoly of the market . . .'. The last bit especially is good for a wry smile, in view of the recent Westinghouse decision to disband its IC division and retire from the market altogether!

I can assure Dr. Eisler that research in a large industrial laboratory, as I know it, is very far removed from the

† *The Radio and Electronic Engineer*, **37**, No. 3, pp. 181–2, March 1969.

picture of a highly organized, ruthless pursuit of monopolistic profit opportunities, so dear to popular imagination. I am amazed that anyone with any experience of the haphazard, groping-in-the-dark, intuitive 'leadership' characteristic of any laboratory where new concepts are developed, could become the victim of such a pathetic fallacy.

The facts are these: no one in Westinghouse, or, for that matter, anywhere else, has any clear idea what to do with paper or flexible transistors. No grandiose plan to corner the world production in paper transistors exists, or is conceivable. All that has happened is that we have made an interesting discovery, with a still unknown potential, which we have discussed in public, in accordance with the usual practice of most large industrial laboratories of releasing such information at an early development stage.

However, one thing I am quite sure about is that we have not set out deliberately to develop a mass-production process suited only for the monopolistic giants of industry, and we are not now proposing such a process. For several years, my group has been exploring the electronics potential of thin-film devices and phenomena. We received no instructions from above, except for the occasional 'why don't you boys do something useful for a change' admonition so familiar to all industrial researchers. After the usual number of false starts, agonizing reappraisals and long lean periods, we have come up with something potentially 'useful'. Had Westinghouse research been controlled by Dr. Eisler's master planners, we would have been planned out of existence many years ago.

The flexible tape thin-film process is admittedly well suited to mass production as it is fast and completely automatable. But so is printing, yet the giant newspaper presses do not threaten the existence of the small printer who specializes in wedding announcements. In much the same way, since the capital investment needed for producing circuits by our process is minute, I can easily visualize the very situation developing which Dr. Eisler finds so desirable, namely, the appearance of many small producers satisfying the demands for specialized circuitry. Our economic calculations show that 'flexible' circuits will be cheaper than either discrete component or integrated equivalents in quantities above a few hundred. We do not, of course, claim that the process, as it now stands, is capable of satisfying all performance requirements of present-day electronic equipment—we are very far from it in many areas, and in others (e.g. in high-speed digital and microwave solid state applications) we are not even trying. As against that, we have some rather intriguing new possibilities which I will not describe here.

When and how Westinghouse is going to utilize our process is an entirely open question. However, I believe that the mass-production approach is the least likely one to be adopted by us, although it may very well be the route taken by some other company. Whether in ten years' time one-half of all circuits in use will be produced by a process similar to ours, or the very concept will have been completely forgotten—who can tell? I share Dr. Eisler's hopes that electronic circuits, with their marvellous versatility of functions, will one day be as cheaply and conveniently produced as printed matter is produced today. I will be satisfied if it turns out that we have made a contribution to this end, and more than satisfied, if, by a combination of luck, insight, persistence and drive on the part of a large number of people involved, the company I work for manages to find the elusive path between our laboratory novelty and eventual commercial success. For all I know, this may lie in the direction of supplying production equipment to large numbers of small users, rather than going into production ourselves. I would be interested in Dr. Eisler's reaction to such a development.

Finally, I would like to say that, despite our disagreement over details, we share Dr. Eisler's vision, outlined in his book many years ago. It may well turn out that the process which ultimately emerges will be a combination of subtractive and additive techniques, and we are in fact thinking along these lines. In any case, I feel that we are on the same side of the fence.

T. P. BRODY

Westinghouse Electric Corporation,
Research and Development Center,
Churchill Boro.,
Pittsburgh, Pa. 15235, U.S.A. *30th June 1969.*

After having given Dr. Brody due credit for his achievements I have criticized his methodological approach from the viewpoint of the potential inherent in the printed circuit technique. I strictly qualified my criticism by pointing out that Dr. Brody's production method is in my opinion a cardinal error of methodological approach *if* his method aims at the highest goal of the printed circuit technique. This qualification is essential for the case I made and this, I trust, has been made quite clear in my letter.

I am sorry that Dr. Brody, in his comments on my letter, felt it necessary to distort my arguments so grossly and to ascribe to me significant expressions which I have not used as characteristic of my views. I have, for instance, neither stated nor based my views on not 'somewhat betraying the printing tradition'—whatever that may mean—nor did I postulate 'universal' laminates containing layers of all materials used in integrated circuits.

My letter referred to my paper 'Printed circuits, some general principles and applications of the foil technique'† and to my book 'The Technology of Printed Circuits' to explain the fundamental technical principle for the choice of method. The principle postulates that 'the printed circuit must in all important aspects be at least as good as the best example of the electrical element it replaces'.

There are numerous important aspects for each class of element. The paper, for example, enumerates those which may be required for the conductors and which are additional to the geometrical configuration produced by the printing and etching process. It deals with: Composition, Thickness, Uniformity, Grain Orientation, Temper, Pliability, Freedom from Pinholes, Strength of Bond to substrate at specific temperatures, Soldering, Welding or other joining characteristics, Surface Roughness, etc. A selection of these characteristics or a variation

† *J. Brit. Instn Radio Engrs*, **13**, pp. 523–41, November 1953.

with various amendments may form part of the specification for any of the elements incorporated in the printed circuit and the problem arises *how to control such numerous characteristics* in production.

If the circuit producer has so many tasks: the control of thickness, of composition, of conformity, etc. *in addition* to the control of the geometrical configuration (shape) he can never be 100% successful in any and all of them. Let us equate the degree of his success in controlling the shape to $X\%$ and that in controlling any other factor to $Y_1\%$, $Y_2\%$, $Y_3\%$... whereby the values of the various Y are suitably weighted according to the importance of the respective characteristic, then his result in a positive process will be proportional to $X\% \times Y\%$, $Y\%$ being the product of $Y_1\% \times Y_2\% \times Y_3\% \ldots \times Y_n\%$.

In the negative method the tasks of control other than the control of the shape ($X\%$) are eliminated. We start with complete laminates and can fully know all the characteristics of the layers before we even design the shape of the elements. We thus deal with definite given characteristics as we have them already, not with some to be attained as near as we possibly can. The success of our negative production method will therefore be proportional to $X\%$ which is much greater than that of the positive method which must be $X\% \times Y\%$. ($X\% > X\% \times Y\%$).

The paper also calls attention to the fact that the laminate used by the negative method described is produced by the specialist techniques found best for the product without any restriction of choice of technique. Any positive method is more limited in both technique and processing, and interfered with by the need simultaneously to produce the layer in a desired precise geometrical configuration. For example, the copperfoil pattern in a usual printed circuit cannot be beaten for bond strength and some other qualities by any deposited copper pattern because the foil and laminate manufacturer had at his disposal a superior specialist technique not usable for the positive circuit producer. The latter's technique is usually a compromise between the different requirements of the various elements with the usually inferior result as consequence.

'Betrayal of printing tradition' is a rather strange summary of a criticism showing up the inferior control and restriction of specialist techniques inherent in any positive method.

I proceed to the concept of 'universal' laminates which are to contain all layers for all the elements in all possible circuits which Dr. Brody quite incorrectly ascribes to me. I do agree with him that such laminates are an unattainable and unfruitful concept, but it is not my concept. From the references given in my letter it can be seen that I have visualized, invented and developed production methods using multi-layer laminates which can be joined to form the most complex circuits. The development of joining methods will determine the limits of the universality of these multi-layer laminates. Anyhow, the number of circuits which can be formed of even a few dozen types of multi-layer laminates is as infinite as the number of words formed from the letters of the alphabet. The analogy with the recording of language in print illustrates the fundamental difference of our respective methods.

Dr. Brody, conscious of the availability of a tremendous range of materials, selects and controls them in his flexible circuit production equipment like the Chinese scholars use their ideographs. They have a choice of many thousand ideographs to write for the reading public, while they can select from several ten thousand ideographs to write for the properly educated. Their record of language has certainly its advantages but it has blocked the cultural development through the printed language perhaps because of the almost 'unlimited reservoir' of ideographs to control.

I have been inspired by the use of the alphabet of about two dozen letters which we join into words. This record of the language is easily read by almost the whole of the population and has been a major instrument in the cultural development of mankind. The analogy with the written language by ideographs and the alphabet shows that while there is the problem of numbers with the ideographs it does not exist for the infinite number of words which can be composed of a little more than two dozen letters of the alphabet. It indicates how I have avoided the problem which Dr. Brody sees in my negative method and it emphasizes my criticism of his positive method *if* he aims at the same goal.

Now to Dr. Brody's spirited defence of Westinghouse and the 'unpolitical' character of such a major technological development as is inherent in printed circuits. I can here cut my comments to this point short by referring him to any professional analysis of the mechanism of large organizations, for instance the works of Professor Galbraith of Harvard. The areas of freedom which are granted to industrial research fit in with the well-publicized image of the big organizations and Dr. Brody may be lucky that Westinghouse keeps his team for brainstorming and playing about, but I find it hard to believe that Westinghouse has no purposeful policy of research and development in the major fields of its activity. The conclusions I have drawn from experiencing the policies of the giant organizations in the U.S. and in Europe are described in my book referred to in my letter. What has happened since, including Dr. Brody's achievement which I went out to acknowledge fully, has not been at variance with my analysis.

PAUL EISLER

57 Exeter Road,
London, N.W.2.

24th September 1969.

U.D.C. 621.374.3:681.32

Signal Processing and Computation using Pulse-rate Techniques

By

J. D. MARTIN,

M.Sc.(Eng.), C.Eng., M.I.E.E.†

Summary: Pulse-rate signals are introduced and the influence of various conventional circuits on their behaviour is considered. An introduction to more complex systems is given by considering the behaviour of certain simple feedback circuits. A survey is then made of the major applications of pulse-rate signals, including regular and stochastic rates, sampled systems, and the digital-differential-analyser.

The aim is to show the common features in the various systems, and to provide a common basis for analysing them. Pulse-rate systems can carry out the normal arithmetic and integral operations. They are characterized by relatively low speed, low accuracy in general, and low complexity.

1. Introduction

A pulse-rate signal consists of a train of pulses whose mean frequency or rate is proportional to the quantity being represented. The pulses themselves are not important: the signal quantity is represented by the rate at which they occur, rather than by electrical signal characteristics.

Pulse-rate signals originate typically from digital tacho devices and variable frequency oscillators. Pulse-rate operations have been used in the past for frequency synthesis[1] and for frequency measurements,[2] but it would be desirable to employ such techniques for many types of control loop where the transducer outputs are of pulse-rate form, e.g. motor speed and acceleration control. Although pulse-rate processing requires relatively complex logic arrangements, the advent of integrated-circuits and the future possibility of large-scale integration make the use of pulse-rate processes a more attractive economic proposition than has been the case so far.[3]

A recent development in digital processors has been the stochastic computer[16] where the processed variable is represented by the probability of a pulse occurrence in a random sequence of pulses. This technique introduces a general class of device, of which the pulse-rate system as mentioned above is a particular case. The stochastic devices enable functions such as multiplication to be carried out very easily, and would therefore lend themselves to adaptive systems where algorithm parameters may be changed as a consequence of previous calculations.

In addition to the simpler types of application envisaged above, it has been suggested that there may be scope for such pulse-rate systems in large processors.[16] Conventional digital processors operate sequentially upon a set of parameters, and the central facilities are time-shared. Consequently there are a number of computational tasks requiring large amounts of complex calculation in real-time, which cannot be served by the conventional digital machine. Some kind of parallel structure is required, in which a number of calculations proceed simultaneously, trading speed for equipment complexity. This approach indicates the use of operational techniques, which usually employ analogue circuits since digital-code operational units are extremely complex. However even analogue circuits present problems when used in large numbers, and a technique employing digital signals and circuits is preferred. A pulse-rate representation of a signal can be routed, transmitted and processed without degradation, and random-rate signals can be processed with very simple circuits.

In general terms, pulse-rate processing elements have similar accuracy and computing speed to analogue elements, but are much simpler and enable multiplication to be carried out very easily. Accuracy and speed can be interchanged, but not in a proportional manner. A doubling of accuracy can be achieved by a reduction of speed by a factor of four, for a random-rate system.

When a digital-code is used to represent a number, the individual pulses or digits are weighted in significance. For example, the most significant digit of a 10-digit binary word would represent a weight of 512, whereas the least significant digit would represent a weight of only 1. Thus although each digit is represented by a similar signal, each must be handled separately in order to preserve its significance; the result of an unintentional shift by one place, for instance, would be disastrous. In practice, this means that for serial representations complicated timing arrangements must be used and for parallel representations complex gating configurations must be employed. With pulse-rate signals, however, as the important quantity is an averaged parameter, the individual

† School of Electrical Engineering, Bath University of Technology, Claverton Down, Bath.

pulses are of little significance, and the rate-signal may be handled very simply without complicated timing or gating arrangements. It is this property which makes pulse-rate signal processing attractive.

The following discussion sets out the principles of how conventional logical building blocks respond to pulse-rates, and then discusses several practical forms of pulse-rate systems. The object is to present a coherent approach to a number of related techniques which have previously been published separately.

2. Pulse Sequences

The parameters of a pulse sequence will first be discussed so that the processing of such sequences or rates may be calculated. In general, a pulse sequence may be either clocked or unclocked, but it is simpler to deal with clocked sequences as the statistical properties of unclocked sequences are less convenient, and processing devices are more easily realized in clocked form. The interval between successive clock pulses is τ, and signal pulses will always occur at clock times although not necessarily at a constant rate.

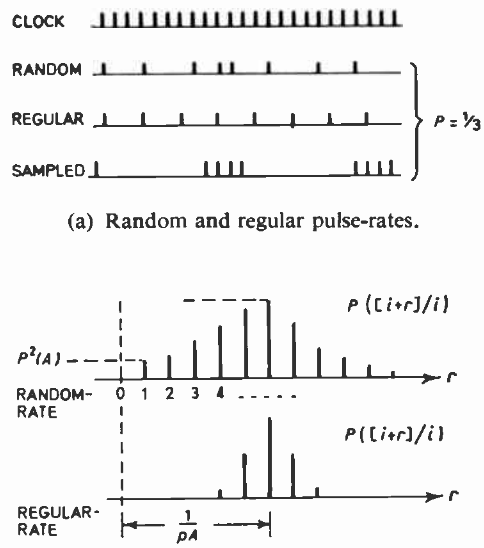
Consider a pulse-rate signal $(A)_r$, where the probability of finding a pulse at a particular clock time is $P(A)$. Then the significance of $P(A)$ is

$$P(A) = \lim_{n \to \infty} \left(\frac{m}{n}\right)$$

where m is the number of pulses actually recorded in an interval consisting of n clock pulses.

Implicit in this definition is the requirement that $P(A)$ is constant for all time, i.e. the process is stationary. There are two particular cases of pulse sequence which must be considered and these form two extremes bounding all practical cases. They are the truly random pulse sequence or stochastic signal, and the regular pulse-rate which represents a single frequency component.

In the stochastic process, successive pulses can occur since the placing of pulses is completely random. In statistical terms, the conditional probability of finding the next pulse r clock periods after the ith pulse in the sequence, $P((i+r)|i)$, is always finite (Fig. 1). However, for the regular rate, although the overall probability of pulse occurrence may be the same, the probability of finding two successive pulses is nearly always zero since the next pulse occurs after an interval of $1/P(A)$ clock periods, which can be predicted. In practice, after processing, regular pulse-rates become somewhat irregular so that it is not possible accurately to predict the position of the next pulse in the sequence; but the dispersion of pulses around their expected positions is never as wide as for the truly random sequence (Fig. 1). It therefore appears that stochastic and regular pulse-rates, although similar in many respects, must be treated differently when considering the details of signal processing.



(a) Random and regular pulse-rates.

(b) Differences in pulse distribution.

Fig. 1. Forms of pulse-rate signal.

When a pulse-rate of either kind is being measured, there is some uncertainty as to the correct value, since the measurement time must always be finite. The stochastic pulse-rate measurement follows a binomial distribution when the average pulse count is being determined, giving a standard deviation[16] $\sigma = \sqrt{nP(1-P)}$ for n clock periods of a pulse-rate having pulse probability P. The uncertainty in measurement may be expressed as $\varepsilon = \sigma/n$ which is standard deviation compared to full-scale magnitude. Thus

$$\varepsilon = \sqrt{\frac{(1-P)P}{n}} \qquad \text{......(1)}$$

For a given value of n, this uncertainty is a maximum for $P = \frac{1}{2}$, and Fig. 2 shows a plot of ε for typical values of n. Although the uncertainty is less for the cases where $P \neq \frac{1}{2}$, clearly the maximum must be taken when comparing the stochastic signal with any other. Considering now the regular rate, the uncertainty is always $+1$ pulse with ideal circuit components. Expressing this too as a ratio, $\varepsilon = 1/n$. Figure 2 also shows this plot.

Examination of Fig. 2 shows that for a given measurement or sampling period, the regular pulse-rate signal gives a more accurate result than the stochastic or random pulse-rate. In practice the signal represented by the pulse-rate is varying with time, and the sampling period is therefore limited to about one

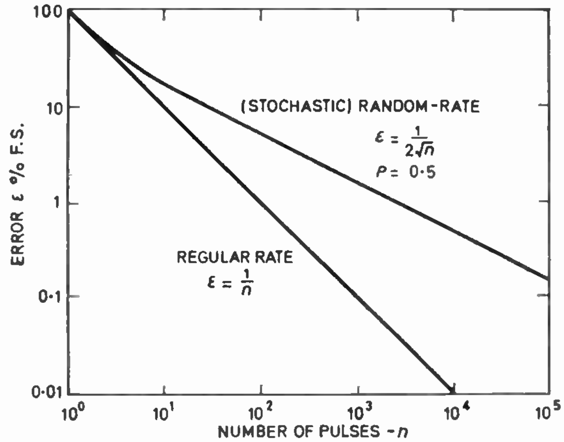


Fig. 2. Error as a function of number of measured pulses.

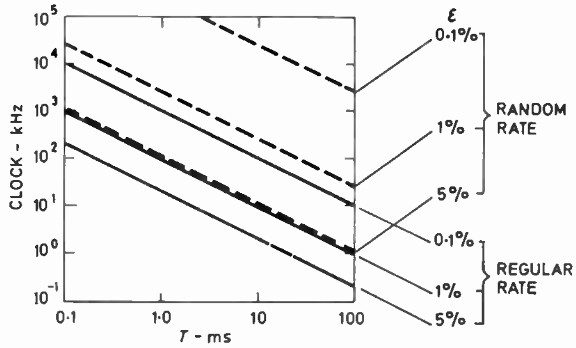


Fig. 3. Clock rate variation with sampling time.

tenth of the signal period. Greater precision may only be obtained by increasing the system clock frequency. Figure 3 shows the trend of system clock frequency versus sampling time, for several degrees of precision, and it is evident that the random rates are penalized at higher levels of precision. In practice however, the beneficial parameters of the regular pulse-rate system cannot be realized, since after processing the pulse-rate possesses some irregularity which is difficult to specify. Practical clock frequencies will therefore lie somewhere between the two extremes for nominally regular pulse-rates.

Random pulse sequences may be readily generated by linear shift-register circuits, which produce maximum-length or m-sequences. This technique provides a good approximation to truly random operation, provided that the repetition period of the pseudo-random sequence is much greater than the sampling period of the system. A shift-register of N stages produces an m-sequence of (2^N-1) pulses, so that a 40-stage register can provide a random signal for $\sim 10^{12}$ clock periods (Fig. 4). A further facility offered by the pseudo-random binary sequences (p.r.b.s.) is that time shifted versions of the sequence have no correlation with one another, and may be generated in principle by adding together pairs of sequences having different delays. This property is very necessary since it will be shown that many uncorrelated random signals are required in a processing system of any size.[17]

The pulse sequences discussed so far have been of unlimited length, and represent a continuous variable. It is possible to conceive a pulse sequence of limited length which therefore represents a sampled-variable. A sequence of this kind consists of a repeated frame of N clock periods, the first M of which contain signal pulses. The variable is thus of magnitude M/N, and has a quantizing interval of $1/N$. This form of operation has advantages when a system is being realized (Sect. 7.3), and does largely eliminate the fluctuations in average rate which are inherent in the continuous pulse-rate representation. Formation of errors will occur as in the continuous pulse-rate system, and Figs. 2 and 3 show that increased precision is gained with longer sampling times.

It is convenient at this stage to define the signal conventions adopted in the subsequent work. A non-dimensional scaler signal which may be a function of time is written as $a(t)$ or just a. The corresponding digital binary signal which represents this basic quantity will be written as A. In this form, the signal representation may be handled by normal switching algebraic forms. When the signal exists in a parallel digital-code form, it will be written $(A)_c$, where $(A)_c < 1$ and a 2's complement convention is implied for signed numbers. When the basic quantity a is represented by a regular rate signal, this rate signal will be written $(A)_r$, and correspondingly written $(A)_s$ for a

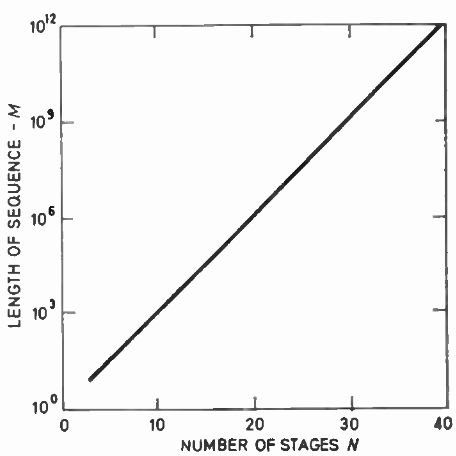


Fig. 4. Pseudo-random sequence length.

stochastic-rate representation. The rate signal (A) is made up of $P(A).f$, where $P(A)$ is the probability variable of pulse occurrence and f is the system clock rate. A stochastic rate is expressed similarly with the difference that individual pulses have a random placing. Signed rate signals may be in one of several forms, which will be discussed in Section 6.

On diagrams, serial signals will be shown by normal single lines and arrows, while parallel signals will be shown as broad arrows with single connecting lines. See Fig. 6 for an example.

3. Operations on Pulse Sequences

Applying pulse sequences of either type to many conventional digital circuits produces novel results, some of which are useful. The following list is not exhaustive, but does represent the most important cases, and will indicate the principles which may be followed. Ribeiro quotes a more complete list.[16]

3.1. AND *Gate*

An AND gate has logical binary inputs A, B and an output C. In switching algebra terms, C = A. B. Suppose now that the inputs are pulse sequences, then the probability of the output C receiving a pulse at any clock time is $P(\mathrm{C}) = P(\mathrm{AB})$: a joint probability of two events. Expressing this joint probability in terms of the individual probabilities, $P(\mathrm{C}) = P(\mathrm{A}).P(\mathrm{B}|\mathrm{A})$, where $P(\mathrm{B}|\mathrm{A})$ is the conditional probability that B will occur given the condition that A has already occurred.

For random pulse-rates, sequences A and B are independent and consequently $P(\mathrm{B}|\mathrm{A}) = P(\mathrm{B})$. The output of the gate is therefore $P(\mathrm{C}) = P(\mathrm{A}).P(\mathrm{B})$, forming the product of the two signals represented by the input pulse-rates, i.e. $(\mathrm{C})_s = (\mathrm{A})_s.(\mathrm{B})_s$. This property is perhaps the most useful and outstanding to be found among simple digital circuits with random pulse sequence signals.

For the case of A and B being regular pulse-rates, the two signals are not entirely independent and the simple product result does not apply. $P(\mathrm{B}|\mathrm{A})$ is difficult to calculate for the general case. Some particular cases can be found where the product rule applies, but these are limited to the type where the periods of the two input rates are integral numbers of clock periods and have no common factors. Multiplication will take place if one of the two inputs is regular and the other random, since this is a condition for independence.

3.2. *Inverter*

An inverter with input A and output B has the logical equation $\mathrm{B} = \overline{\mathrm{A}}$. At clock time, the probability equation is $P(\mathrm{A})+P(\mathrm{B}) = 1$, so that $P(\mathrm{B}) = 1-P(\mathrm{A})$. This output signal must be clocked, and clock interrogation may take place after the gate, or alternatively the gate may be considered to be normally inhibitory but enabled by the clock. This is equivalent to regarding the gate as an AND gate with inputs of mixed sense, so that the output probability is multiplied by $P(\mathrm{C})$ the clock probability, which is unity. In terms of the rate signals, $(\mathrm{B})_s = 1-(\mathrm{A})_s$.

Provided that the pulse-rate is expressed as a fraction of the maximum, this operation is also valid for regular pulse-rates.

3.3 *Squarer*

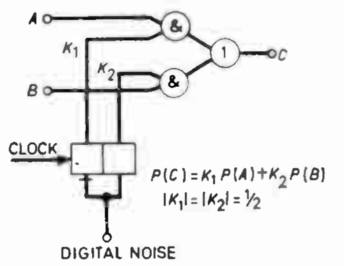
Suppose that the AND gate above has the same signal applied to both inputs in an effort to obtain $P^2(\mathrm{A})$. Under these conditions $P(\mathrm{B}|\mathrm{A}) = 1$ and the output is in fact $P(\mathrm{A})$. Input B must have a signal whose probability of occurrence is $P(\mathrm{A})$, but which is independent of, or uncorrelated with A. This condition is satisfied if the pulse sequence applied to input B is delayed by one clock period. The AND gate now gives an output pulse for two successive input pulses in sequence A, and the probability of this event is $P^2(\mathrm{A})$. This philosophy may be extended to higher power laws by providing more gate inputs, but an additional delay element must be provided for each additional input.

Because of its statistical nature, this operation is only valid for stochastic pulse-rates.

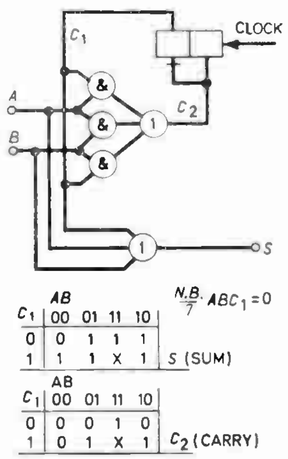
3.4. OR *Gate*

An OR gate with inputs A, B and output C has the relation C = A+B. The probability equation is $P(\mathrm{C}) = P(\mathrm{A})+P(\mathrm{B})-P(\mathrm{AB})$, since two coincident inputs are registered as only one input and a pulse is lost. Expanding, the joint probability $P(\mathrm{AB}) = P(\mathrm{A}).P(\mathrm{B}|\mathrm{A})$. For this occasion, a useful result will be obtained if $P(\mathrm{B}|\mathrm{A}) = 0$, which requires the two inputs to be dependent but exclusive. The inputs to the gate should therefore be gated by a pair of exclusive random signals, $\mathrm{K}_1, \mathrm{K}_2; P(\mathrm{K}_1\mathrm{K}_2) = 0$. These two signals may be formed by a local flip-flop element (Fig. 5(a)), and the final relationship is $P(\mathrm{C}) = P(\mathrm{K}_1).P(\mathrm{A})+P(\mathrm{K}_2).P(\mathrm{B})$. The maximum value of $|P(\mathrm{K}_1)+P(\mathrm{K}_2)| = 1$.

Addition of regular pulse-rates is also possible, but by a different approach. As mentioned above, the condition which violates the additive property of the OR gate is that of coincident input pulses. If the probability of two successive pulses is zero for both input pulse-rates, then the lost pulse may be stored after the style of a carry digit and added at the next clock time (Fig. 5(b)). Notice that if two regular pulse-rate signals are added in this way, the sum will not be strictly regular. At best it will have periodic fluctuations determined by the difference frequency of



(a) Stochastic.



(b) Regular.

Fig. 5. Pulse-rate adders.

the two input rates, and this will be further disturbed by the irregularity caused by coincident pulses.

3.5. *Counter or Accumulator*

The counter is a special case of an accumulator, adding only a count of unity at each operation, whereas the accumulator may allow the addition of a variable number greater than unity at each input pulse. Consider a counter with an input $(X)_r$, counting a total of N input pulses in T seconds, where the ith pulse has a period t_i. Then the average period of the input signal is

$$\frac{1}{N}\sum_{i=1}^{N} t_i = \frac{1}{\text{av}\,(X)_r} \qquad \ldots\ldots(2)$$

where av denotes average. Then the count displayed by the counter will be

$$\frac{T.N}{\sum_{i=1}^{N} t_i} = T.\text{av}\,(X)_r = \int_0^T (X)_r.\text{d}t \qquad \ldots\ldots(3)$$

Thus the digital-code signal $(A)_c$, which forms the counter output, is:

$$(A)_c = K^{-1}\int_0^T (X)_r.\text{d}t \qquad \ldots\ldots(4)$$

where K is the full-scale counter output, and $(A)_c < 1$.

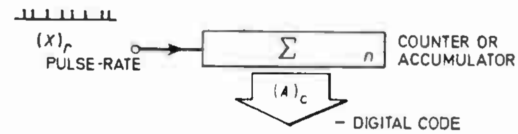


Fig. 6. Pulse-rate integrator.

Thus a counter forms the time integral of the input pulse-rate, presenting it as a digital-code signal, and this is true for regular and stochastic pulse-rates (Fig. 6). This property may be extremely useful, since the integral can be accumulated or held over an indefinitely long period without degradation. The integral value is of course quantized to ± 1 count, i.e. the true value relative to the indicated value is

$$(A)_c\left\{1 \pm \frac{1}{(A)}\right\}.$$

This error may be made as small as desired by increasing the actual count.

There are several improvements which may be made to the basic counter. Consider its use as an accumulator, where at each input pulse a number $(B)_c$ is added into the accumulator. The integral is then scaled up by this factor. For full operational use a counter must be of the reversible kind and an accumulator must be capable of subtraction. If full flexibility is required, the counter must also be capable of counting through zero into the region of negative count, but this facility may be easily added. An initial condition may be set into the counter or accumulator before integration starts.

It is also necessary, for full operational use, to derive a rate output from the counter to form the input signal to the next stage. Section 4.1 describes code/rate converters for both regular and stochastic pulse-rates, and these may be used as appropriate.

One technique for taking an output is to provide an output pulse every time the counter or accumulator overflows. Thus, if overflow occurs when the count increases by an integer δa, and the interval between successive overflow pulses is given by δt, the rate-of-change of count is

$$\frac{\text{d}a}{\text{d}t} = \frac{\delta a}{\delta t}$$

The mean output rate is therefore

$$(Z)_r = \text{av}\left(\frac{1}{\delta t}\right)$$

which is numerically

$$\frac{1}{\delta a}\left(\text{av}\,\frac{\text{d}a}{\text{d}t}\right)$$

But since $a = \int x.\text{d}t$, the output rate

$$(Z)_r = \frac{\text{av}\,(X)_r}{\delta a} \qquad \ldots\ldots(5)$$

This form of output is useful in d.d.a. applications (Sect. 7.4), and although not affecting the mean rate of $(X)_r$ except by a constant factor, does operate as a smoothing operator for irregular pulse sequences.

The smoothing action of the circuit can be seen without performing a rigorous analysis of the statistical properties of the output rate. Suppose the probability of a pulse occurrence is p, then for a truly stochastic input rate, the probability of r successive pulses is p^r. Taking the example cited above, an output pulse will be generated after δa input pulses. Thus the probability of r output pulses at the minimum interval of δa, is $p^{r.\delta a}$. The result of this operation is therefore to reduce the bunching of the input pulses, relative to the new reduced output rate.

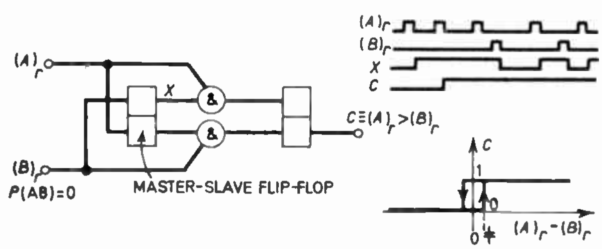


Fig. 7. Pulse-rate comparator.

3.6. *Comparator*

Comparison of two pulse-rates is equivalent to subtracting one from the other and examining the sign of the resultant. Subtraction of two regular pulse-rates may be carried out by a circuit of the form in Fig. 7. For this circuit to give the correct output at all times, there must never be more than one B pulse between any two A pulses when $P(\text{A}) > P(\text{B})$. If the rates are more irregular than this, then a multi-state counter must replace the single flip-flop so that n B pulses may be stored between two A pulses, and the A pulse sequence debited with this account before a definite change of sign is recorded. In fact this operation involves a sampling period, which is necessary before any decision may be made. Arithmetic operations may be carried out pulse by pulse, but when a measurement of pulse-rates takes place, a finite sampling time is inherently necessary.

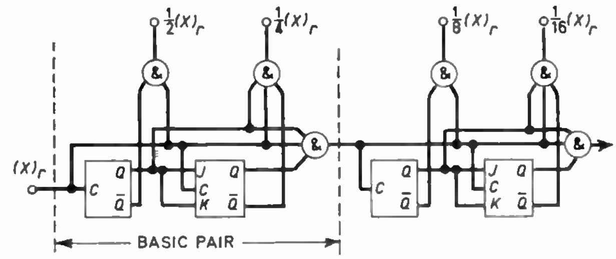


Fig. 8. Binary scalar.

Other modifications may be made, and such comparators may be adapted to form several interesting structures for the accurate comparison of regular pulse-rates.[4]

3.7. *Binary-rate-multiplier*

The binary-rate-multiplier (b.r.m.) is a well-known device, and has been described by Wood,[2] Lundh[5] and Crabbe,[6] among others. The principle of operation is an illustration of the pulse-rate processing techniques discussed in the previous sections.

The basis of this device is a binary scalar which divides down the input pulse-rate, giving pulse outputs rather than the square-wave outputs of a normal binary counter. Figure 8 shows the structure of a typical scalar employing d.c. coupling and semi-parallel drive.

The output of the ith binary scalar stage is $\{2^{-i}(X)_r\}$, and this is gated by the $(n-i)$th digit of the input code $(A)_c$, which has n digits. The scaler outputs are arranged to be non-coincident, so that rate addition may take place in the output OR gate (Fig. 9). Expressing the input code as

$$(A)_c = \sum_{j=1}^{n} a_j.2^{-j}$$

the output rate is seen to be:

$$(B)_r = \sum_{i=1}^{n} a_i.2^{-i}(X)_r$$
$$= (A)_c.(X)_r \qquad \text{where} \qquad (B)_r < (X)_r \qquad \ldots\ldots(6)$$

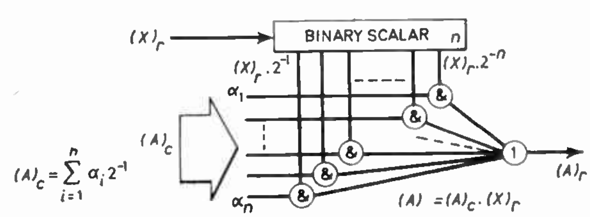


Fig. 9. Binary rate multiplier.

This output rate is not regular, since the ith scaler stage contributes a rate of $2^{-i}(X)_r$, and in the general case an irregular pulse sequence results. However, the pattern is cyclic having a period of $2^n/(X)_r$, and this time may be used as a basis for assessing the minimum sampling time.

When average pulse-rates are considered, the b.r.m. is thus a convenient device for effecting multiplication for regular input rates. It is not suitable for stochastic rates since the scaler outputs would be smoothed, as discussed in Section 3.5. The instantaneous fluctuations in output rate when the input rate is regular are not very large, and when the output is to be integrated, as is often the case, the error is even less significant. Yang[7] has estimated this error with the following result.

A b.r.m. is considered having $(n+1)$ stages, and the error is estimated as the difference in a given time between the actual number of pulses at the output and the number of pulses which would correspond to an ideal straight-line transfer characteristic. It is shown that the maximum error, which requires a particular code at the input, is given by:

$$\Delta_M = \tfrac{1}{6}[n+\tfrac{1}{3}\{10+(-1)^{r+1}.2^{-n}\}] \qquad \ldots\ldots(7)$$

where $r = 1$ for n odd, and $r = 0$ for n even.

For a realistic value of n, the maximum error appears to be

$$\Delta_M = \tfrac{1}{6}\{n+3{\cdot}33\} \qquad \ldots\ldots(8)$$

Thus the error, which is normally quite small, may be reduced as required by increasing the number of b.r.m. stages. This is reasonable, since the stages of lower significance interpolate between the coarse quantizing steps of the stages of higher significance. Notice that this error is an absolute quantity, and the relative error may be more significant in certain cases. Yang also gives the code inputs under which this maximum error occurs, so that a more detailed assessment of the error position may be made. In general this degree of complication is not necessary.

The binary-rate-multiplier is suitable for regular rate applications where an average measure of output rate is satisfactory, although the instantaneous output rate may show quite large fluctuations. The effect of this latter property will be dealt with in Section 5. Notice that a pure binary code has been tacitly assumed for convenience, but there is no reason why a b.c.d. code weighting should not be used.

3.8. *Binary-rate-divider*

This device is an often-used combination of binary counter and binary number comparator (Fig. 10), but forms a useful tool for pulse-rate processing. The counter is allowed to accumulate pulses from the input rate $(X)_r$, until the total is equal to the scaled input code $(A)_c$, when the counter is reset and an output pulse generated. Thus for a regular rate input, and a constant input code, the output rate is

$$2^{-n}\frac{(X)_r}{(A)_c}$$

Since this operation inherently involves division, it is potentially useful and in fact fairly unique in signal processing circuits of any kind. It will be shown that a

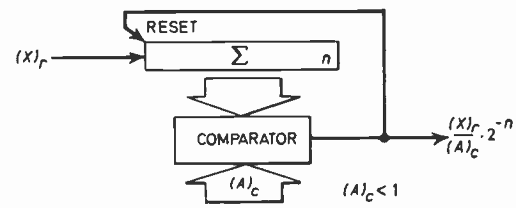


Fig. 10. Binary rate divider.

kind of duality principle can be invoked, where the b.r.d. may be interchanged with a b.r.m. in certain cases, to provide a similar end result but with different characteristics. Since the division process is a fundamental part of its behaviour, the b.r.d. does not introduce any fluctuations into the output signal, but in fact operates to smooth any fluctuations in the input signal as shown in Section 3.5. It is thus not directly suitable for stochastic rate inputs.

4. Signal Converters

The utility of any signal form is largely determined by the ease with which conversion may be made into and out of other signal forms. Thus the regular pulse-rate and stochastic pulse-rate signals must be readily convertible with analogue and digital-code signals.

Converters fall into two groups, those that are direct and those that involve a feedback loop containing a converter of the complementary type. Each converter has output and input signals, and a reference signal which although fixed may be varied according to some third variable to yield a multiplying or dividing converter. Such operation is sometimes very useful.

When conversion between two different forms of signal is considered, it generally occurs that conversion in one direction is straightforward and direct, while conversion in the opposite direction must be carried out by a feedback configuration. Feedback type circuits will be considered separately in the next section.

4.1. *Code/rate Conversions*

Code-rate converters (c.r.c.) are of the direct type and differ in form according to whether a regular or stochastic rate is required. The stochastic rate device is shown in Fig. 11, where the input digital code is compared with the output of a random number generator.[17] Random numbers may be produced by sampling a cycling counter, or by using a pseudo-random sequence. An output pulse is produced at clock time when the input code $(A)_c$ is greater than the random number $(R)_c$. Suppose that there are M possible random numbers, then if they are all equally

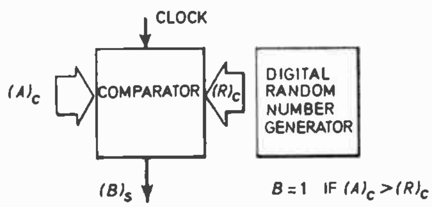


Fig. 11. Stochastic code/rate converter.

probable, the probability of any one occurring is $1/M$. The probability of an output pulse occurring is then

$$P(B) = 2^n \frac{(A)_c}{M}$$

for a binary number system of n digits ($(A)_c < 1$); and a linear conversion is achieved. Notice that if not all the random numbers are equally probable, then a non-linear conversion law will result. The scale factor depends on the total repertoire of random numbers M, and is not easily modified.

Taylor[18] has described a means whereby a number of independent stochastic rate signals may be derived from one basic generator employing two feedback shift registers. These signals could therefore form the inputs to a stochastic rate system, converting input digital-code signals into stochastic rates.

The b.r.m. may be used for code/rate converters of the regular rate type, although as mentioned previously some rate irregularity is present. The b.r.d. may also be used, to produce rates having an inverse relationship to the input code.

An alternative type of code/regular-rate converter has been proposed by Butaev and Romashkan,[8] which makes use of the overflow of a counter. Using the symbols introduced previously rather than those in the mentioned paper, the circuit has an input digital-code $(A)_c$ having p digits, and an input rate $(X)_r$. A parallel adder adds the code $(A)_c$ into an n-stage accumulator at every pulse in $(X)_r$, and when overflow occurs the accumulator is reset. The accumulator overflows when $m(A)_c > 1$ where m is the number of pulse of $(X)_r$ since the last accumulator reset. It is shown in the paper that the output rate is approximately given by

$$(Z)_r = (X)_r.(A)_c[1+2^{-(n-p)}]^{-1} \qquad \ldots\ldots(9)$$

The transfer characteristic is therefore linear except for the small error term which may be made to approach zero. Notice that this error term is due to the fact that 2^n is not divisible by many values of $(A)_c$. Had a reset not been applied at each overflow, the output rate would show noise type irregularities. For this linearity error to be less than the quantization error, $n > 2p$. Since the action of this circuit depends on accumulation, it is not suitable for stochastic rates owing to the smoothing effect.

4.2. *Rate/code Conversions*

The most widely used rate/code converters have been the feedback type which will be discussed under Section 5. However, direct converters have been suggested by Vincent and Rowles,[9] and Martin.[4]

The Vincent and Rowles converter consists basically of a counter which accumulates according to the input pulse-rate. However at intervals of time T_s, a fraction of 2^{-n} of the count is subtracted from the accumulated total. The arrangement therefore behaves as a sampled system with exponential response, having a time-constant $T_c = 2^n T_s$. The hardware performs subtraction by parallel-adding the 1's complement and producing an end-around-carry. This converter will handle stochastic rates as well, operating on them after the style of a first-order lag filter.

The Martin converter relies upon the condition that the input rate is regular, and uses a cascade of comparators like that in Section 3.6 in conjunction with the outputs of a binary scalar, to produce a direct comparison between the input rate and the reference rate driving the scalar.

4.3. *Rate-analogue Converter*

Rate-analogue converters (r.a.c.) are usually of the direct type. The simplest consists of a pulse regenerator producing defined-energy pulses, and followed by a low pass filter. The only difference between a regular-rate and a stochastic-rate converter is the required filter bandwidth for a given output uncertainty. (See Fig. 2.)

4.4. *Analogue-rate Converter*

Random-rate converters may be produced by a method similar to that in Fig. 12.[16] Noise is mixed

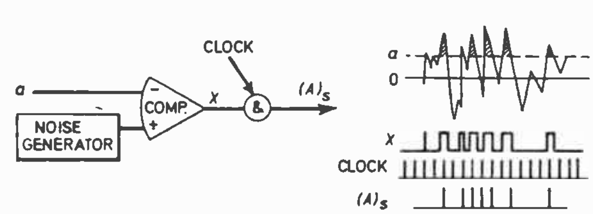
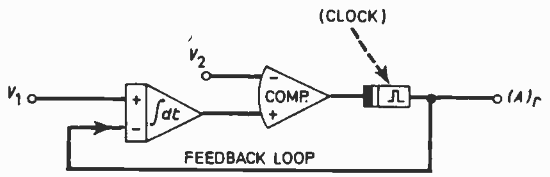


Fig. 12. Simple stochastic analogue/rate converter.

with the analogue signal and the resultant clipped to produce binary noise, which is strobed by the system clock. The probability of a clock pulse being propagated is proportional to the total time for which the binary noise waveform is ON, and this is proportional to the analogue signal if the original noise waveform has a rectangular probability density distribution.



MODE	$(A)_r$
OPEN LOOP	$K_0 \frac{V_2}{V_1}$ (COMP. OUTPUT RESETS INTEGRATOR)
CLOSED LOOP	$K_1 V_1 + sK_2 V_2$

Fig. 13. Principle of regular analogue/rate converter.

Many circuits exist for regular-rate a.r.c.s, but that in Fig. 13 will illustrate the general principle quite well. When operated on open-loop, the comparator output resets the integrator and propagates an output pulse which may be clocked or unclocked. The output pulse period is then proportional to V_2/V_1. Thus if V_2 is taken as the input analogue signal and V_1 is kept constant, a rate results which is proportional to $(V_2)^{-1}$. This could conceivably be useful in certain cases, since the basic integrator circuit is probably simpler than the b.r.d. described earlier.

If the loop is closed, then a different mode arises where the output pulse rate is proportional to

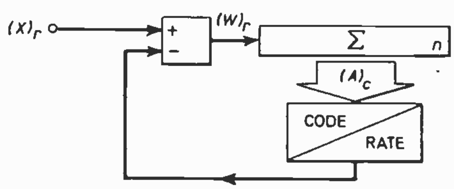
$$\left(K_1 V_1 + K_2 . \frac{dV_2}{dt}\right)$$

Using V_2 as the input variable gives a pulse-rate proportional to input voltage rate-of-change, and is hence directly suitable for input to a digital differential analyser (Section 7.4). This circuit is used in communication applications under the guise of a delta modulator.[10] Use of V_1 as the input variable gives a rate proportional to input, and the circuit is sometimes called a delta-sigma modulator.[11]

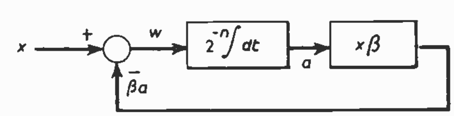
This simple hybrid circuit would seem to be quite versatile and useful as a standard system building block, perhaps suitable for medium-scale integration construction.

5. Feedback Systems

The discussion thus far has concentrated on the behaviour of single circuit blocks when associated with pulse-rate signals. A wide and interesting field is opened up when such blocks are arranged into subsystems, particularly those involving feedback. The basic principles of feedback systems will be illustrated by some particular examples. An approximate analysis of these feedback configurations will be given, assuming that fluctuations in the signal pulse-rates occur over a large number of pulses, thus implying that continuous-signal mathematics may be employed. For a more precise description of the behaviour when this assumption is not valid, sampled-data techniques must be used.



(a) Circuit block diagram.



(b) Signal diagram.

Fig. 14. Binary-rate-multiplier (b.r.m.) type feedback loop.

Figure 14(a) shows a basic feedback loop configuration which consists of a counter or accumulator, whose input is the difference between an input signal and a feedback signal derived from the counter contents. Analysis of the equivalent signal circuit in Fig. 14(b) shows that for a code-rate converter having a proportionality constant β, the signal quantity a represented by the counter contents $(A)_c$, is

$$a(s) = \frac{2^{-n}.x(s)}{s+\beta 2^{-n}} \qquad \text{......(10)}$$

where s is the Laplacian operator.

Consequently, the feedback circuit behaves as a first-order lag, with effective time-constant $2^n/\beta$.

For regular rate signals, the code-rate converter would be a binary-rate-multiplier, with $\beta = (Y)_r$ ($(Y)_r$ being a reference rate). The time response to a suddenly applied constant input rate, would therefore be

$$a(t) = \frac{x}{y}\left[1-\exp\left(-ty2^{-n}\right)\right] \qquad \text{......(11)}$$

The circuit would therefore operate successfully as a continuously-reading frequency meter, when steady conditions were obtained:[2]

$$(A)_c = \frac{(X)_r}{(Y)_r} < 1 \qquad \ldots\ldots(12)$$

For this use, the time-constant $2^n/(Y)_r$ should be large, giving good resolution and a long averaging time.

Although this principle appears quite satisfactory, there is a considerable problem in practice, which arises out of the non-ideal behaviour of the b.r.m. As noted in Section 3.7, the b.r.m. is an average rate device, and individual output pulses are spaced apart by an integral number of reference-rate pulse-periods. Consequently the output pulses are irregular when viewed in pairs, although the actual cumulative error is not very large (Sect. 3.7 and Ref. 4). When used in this feedback circuit, the b.r.m. causes the difference rate as applied to the counter to have large instantaneous fluctuations resulting in irregular sign changes and consequent hunting in the counter itself. This effect has been discussed by Lundh,[5] and cured by using backlash circuits which only allow a sign change at the counter after the instantaneous rate difference exceeds a given threshold. The technique is to use a comparator circuit as described in Section 3.6, which pairs pulses from the two inputs and only produces an output pulse of either sign when excess pulses are received. Cascading several of these comparators will enable the backlash magnitude to be increased by one for each additional backlash circuit. In this way it is possible to stabilize the loop, although as yet there has been little indication as to how many backlash circuits would be required. However, using the maximum b.r.m. error analysis of Yang[7] as discussed in Section 3.7, it would appear that, to ensure stability, the number of backlash units should exceed the maximum instantaneous b.r.m. error. The addition of backlash in this loop does not fundamentally affect the loop behaviour, since the additional delay incurred is only one or two pulse periods, which is negligible compared to the number of periods required for significant definition of pulse-rate.

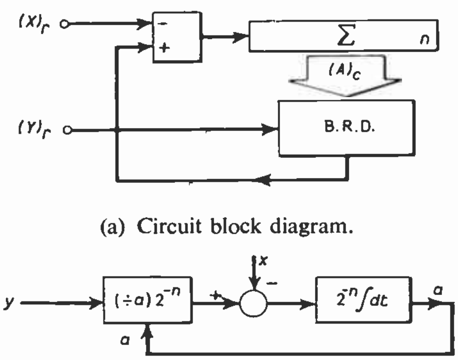
A further potential use of this particular feedback circuit can be seen by considering the difference signal at the counter input $-(W)_r$:

$$w(s) = \frac{2^{-n}s.x(s)}{s+\beta 2^{-n}} \qquad \ldots\ldots(13)$$

This signal corresponds to the approximate rate-of-change of the input signal if the time-constant $2^n/\beta$ is small compared to the time scale of variations in $(X)_r$. The optimum value for $2^n/\beta$ approaches zero, leaving the counter nearly empty under steady conditions which would give rise to considerable hunting due to the coarsely quantized nature of the counter output. Notice too that for correct operation under these conditions, the reversible counter must be capable of running smoothly through zero and affecting the b.r.m. output accordingly. This facility is quite simple to arrange. The b.r.m. feedback loop therefore has a dual role and can be optimized for each duty by control of the number of counter stages and the reference rate $(Y)_r$.

A stochastic version of this type of closed loop may be formed by using a suitable code-rate converter as described in Section 4.1. Similar equations obtain to those above, with a time-constant $M\tau$, where τ is the clock period. Since the pulse-rates are stochastic anyway, there is no trouble with local irregularities as in the case of the b.r.m. and regular rates. This device is used as an output unit in certain stochastic systems, where it has been called an 'addie'.[17]

An alternative form of feedback circuit having the same frequency-meter type of performance as the



(a) Circuit block diagram.

(b) Signal diagram.

Fig. 15. Binary-rate-divider (b.r.d.) type feedback loop.

b.r.m. loop in Fig. 14 may be formed by using a b.r.d. as a feedback element (Fig. 15(a)). Taking rate input $(X)_r$ as a constant reference rate, the loop differential equation may be set up as in Fig. 15(b):

$$2^n . a\frac{da}{dt} + xa = 2^{-n}y \qquad a < 1 \qquad \ldots\ldots(14)$$

Thus the steady-state response of the loop will be

$$(A)_c = \frac{2^{-n}(Y)_r}{(X)_r} \qquad \ldots\ldots(15)$$

Although the steady-state response is of the same form as that of the b.r.m. loop, eqn. (14) indicates that the transient response will be altogether different. Taking the simple case of a step input of magnitude $(Y)_r$, and an initial condition of $(A)_c = 0$, the solution of this equation is

$$2^n\left[\frac{2^{-n}y}{x^2} . \log_e\left(\frac{2^{-n}y}{2^{-n}y - xa}\right) - \frac{a}{x}\right] = t \qquad \ldots\ldots(16)$$

The time response of this function is not at all clear, but a useful comparison may be made between the b.r.m. and b.r.d. loops for the case where the input is a full-scale step.

Considering this case for the b.r.d. loop, the maximum counter indication is $a_m = (1-2^{-n})$, giving the input signal magnitude as $y_m = 2^n.x.(1-2^{-n})$. Now in order to estimate the practical response, consider the time t_1 during which the counter contents built up to $(1-(1+b)2^{-n})$, i.e. a resolution of

$$\left(\frac{b}{2^n-1}\right)$$

Then

$$t_1 = \frac{2^n}{(X)_r}[0{\cdot}7n - \log_e b - 1] \qquad \ldots\ldots(17)$$

if $2^n \gg 1$.

For the b.r.m. loop, the counter maximum will be the same as for the b.r.d. loop, but the maximum input signal is $x_m = y(1-2^{-n})$. Taking the same measure of resolution, the response time will be:

$$t_1 = \frac{2^n}{(Y)_r}[0{\cdot}7n - \log_e b] \qquad \ldots\ldots(18)$$

Now, comparing the two loops when they have the same maximum input $(W)_r$, the b.r.m. requires a reference of $(W)_r$ while the b.r.d. loop requires a reference of $2^{-n}(W)_r$. The corresponding response times are therefore in the ratio of about 2^n, the b.r.d. loop having the longer response time. The b.r.d. loop does not therefore appear to be a serious competitor to the b.r.m. loop in this context, although it does not suffer from the instantaneous rate fluctuations inherent in the b.r.m.

However, if the loops are compared on the basis of similar reference rates $(Z)_r$, then the input range of the b.r.m. loop is 0 to $(Z)_r$ approximately, while the input range of the b.r.d. loop is approximately $(Z)_r$ to $2^n(Z)_r$. A possible use for this property is the measurement of very high pulse-rates, where only the b.r.d. itself need operate at the input rate and the reference can perhaps be a more convenient value.

The concept of the b.r.d. loop is important since it introduces the idea that a comparison may be made between pulse-rates which are not within the same range. Thus if the b.r.m. loop of Fig. 14 has included within it a b.r.d. with input $(B)_c$, the steady-state response becomes

$$(A)_c = 2^n.\frac{(B)_c(X)_r}{(Y)_r} \qquad (A)_c, (B)_c < 1 \qquad \ldots\ldots(19)$$

Considerable scope is now given for range variation within the limits of $(Y)_r$ and $2^n(Y)_r$, although the loop time-constant is now given by

$$\frac{2^{2n}(B)_c}{(Y)_r}.$$

The b.r.d. loop is not directly applicable to stochastic rate signals, owing to the smoothing property of the b.r.d. itself (Sect. 3.5). A stochastic element has been devised which does enable division to be carried out indirectly.[19]

6. Signal Coding

Rate signals are inherently unipolar, so some form of coding is necessary before bipolar problem variables can be represented. Some of the possible methods of carrying bipolar signal information are well illustrated by considering one or two of the current schemes for stochastic systems. There have been several different coding schemes proposed for stochastic schemes by Ribeiro,[16] but the simpler ones will be discussed here.

The first uses the single-line bipolar rate signal by employing a transformation of the form $x_0 = 2x-1$, where x_0 is the problem variable and x is the probability variable. Thus $x_0 = 0$ is represented by $x = 0{\cdot}5$. This scheme enables the generation of negative numbers by inverters, since $-y_0 = (1-y)$; but multiplication must be carried out by an equivalence gate. Gaines[17] gives a set of elements for processing signals using this code, and also points out that the probability signal variance is a maximum for $x_0 = 0$, which may prove to be a disadvantage. In general however, this is a straightforward method of coding and only uses one connecting wire for each signal path. Although this coding is convenient for random-pulse systems, it raises considerable problems of decoding when applied to straight pulse-rate systems.

The methods found to be most satisfactory for regular pulse-rate schemes are those which employ two-line conventions. In the first, one of the lines carries the pulse-rate which therefore indicates the magnitude of the signal, while the other line carries a sign signal which will normally be continuous. Subtraction, while being satisfactory for regular rates, is inherently difficult in this coding for stochastic rates, since it is necessary to store large numbers of pulses from minuend or subtrahend while further pulses are awaited from the converse channel. The problem is one of comparison as discussed in Section 3.6.

A further two-line coding scheme employs one line to carry pulses when the signal is positive, and the other to operate when the signal is negative. Pulses therefore only occur on one line at a time, and the variable is given by the difference in pulse-rates over a suitable averaging time. Negative quantities may be formed by interchanging the two signal lines, and the other arithmetic functions may be readily formed.[17]

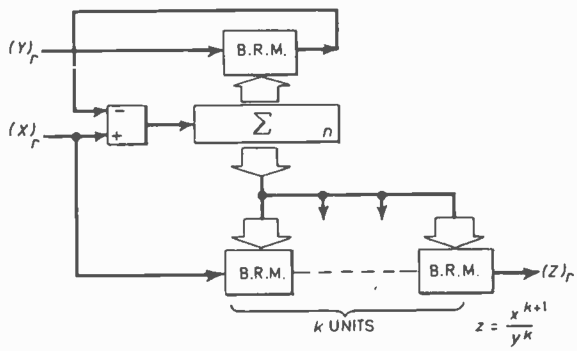


Fig. 16. Generation of integral powers.

In general it is not easy to effect a change between conventions within a stochastic system, but changes between the two-line conventions are easily effected in regular-rate systems by using circuits of the comparator type (Sect. 3.6). None of the various methods of signal coding appear to be particularly universal, so it is best to select the coding for each individual circumstance.

7. Systems and Applications

The techniques discussed above have been used in a number of specific applications, since the ideas are not all recent. Some comments will now be made as to how the basic elements may be built up into several types of system, and some indication of typical applications and limitations will be made. This list is not exhaustive, but will indicate the extent to which these techniques are practical.

7.1. *Regular-rate*

Lundh[5] was one of the first to investigate the possibility of using regular-rate processing to perform conventional computational operations, and Phillips[3] has more recently listed a number of quite complicated functions which may be formed. These complex functions rely on the multiplying property of the b.r.m. to form products, divisions and roots, all of which are difficult in analogue form and clumsy in conventional digital arithmetic. As examples of the kind of functions which may be solved, consider Fig. 16 which shows a circuit to produce the *k*th power of an input, and Fig. 17 which shows a square rooting circuit. The operating equations may be seen by inspection using the already established principles. The dynamic performance of such circuits is difficult to determine in general, although the principles used in Section 5 may be extended to more complex cases provided that the signal fluctuations due to irregularities do not become too large. Lundh does however show how the square root function may be analysed, and also gives an experimentally determined settling curve.[5]

In principle, it seems possible to form integro-differential equations, extending the integrating function of the counter, and perhaps using the time-differentiating properties of the feedback loop (Sect. 5). Care must be exercised in this latter case, due to the degree of approximation involved.

The limitation with algebraic functions of high complexity, is the cumulative irregularities introduced into the otherwise regular pulse-rates by the arithmetic operations. The b.r.m. irregularity has already been noted, but also the addition or subtraction of two completely regular pulse-rates must produce a result which has some irregularity in it. The only means by which this may be reduced is to employ a smoothing circuit of the required time-constant, and this causes delays in computation. However the circuits are suitable for limited parallel processing of a number of signals.

It appears therefore that for computation use, regular-rate systems are likely to be limited to calculations of small complexity, and would find greatest application where the input signals are already in rate form, or where only a small local processor is required. At the 1968 IFAC symposium on pulse-rate methods,[12] it was suggested that the conventional three-term controller still appears to be the most suitable for many process control applications, and a pulse-rate version of this may well prove more satisfactory than the present analogue arrangement.

The pulse-rate system has the great advantage that it interfaces very easily with analogue and digital-code signals, conversion being relatively simple to effect, compared for example with analogue/digital-code conversion (Sect. 4). A number of published applications deal with conversions of this kind,[2, 4, 8, 9] and the pulse-rate does form a convenient transfer or bridging signal in the digital measurement of analogue quantities. A number of digital voltmeters use this principle, and the addition of various pulse-rate units would enable scale changes and conversions to be effected quite easily by a program sequence, thus enabling a single printer to display corrected and scaled indications from a number of widely different sources. The limitation in this example is of course the relatively low speed, but there are a number of

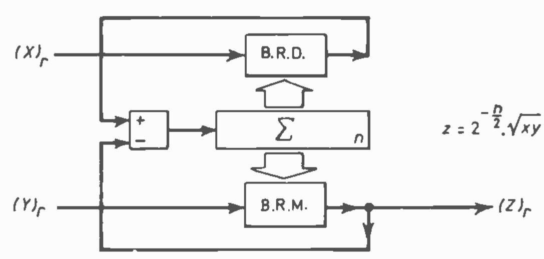


Fig. 17. Square-root generation.

industrial multiplexing and scanning applications where high speed is not essential and this technique could well be used.

There are several measuring transducers which give regular-rate outputs, and these would form natural input sources for a pulse-rate system.[15] The most common types are those involving rotation, where a digital tacho is used to provide a pulse-rate proportional to rotational speed; and there is also a large class of transducers employing variable inductance or capacitance elements which are normally included in an oscillator circuit, producing a variable frequency output dependent on the measured quantity.[15] Pulse-rate techniques could be used here for scaling, limited computation and maybe linearizing of the output.

In the preceding discussion, the important features of quantizing, sampling and round-off errors have been neglected since they form an intricate subject. A considerable amount of work seems to have been published on these aspects at the IFAC symposium mentioned above.[12]

Although at first sight pulse-rate techniques appear ripe for exploitation in the field of frequency synthesis, the resultant rate irregularities in the b.r.m. tend to make it impractical. Rey[1] made an initial study of this problem from the pulse-rate approach, but the conventional frequency synthesizer uses a phase-locked oscillator whose output frequency is divided down by a variable ratio and compared with a fixed reference frequency.[13] It would appear however, that a device such as that described by Butaev and Romashkan,[8] may enable a stable fixed frequency source to synthesize directly a variable output frequency.

7.2. *Stochastic Rate*

The stochastic rate elements are characterized by extremely simple operations of multiplication and addition, namely AND and OR gates respectively in the simplest case. In order to operate a large number of arithmetic steps, a large number of uncorrelated stochastic pulse-rates are needed, which therefore require some additional hardware.[17] However hardware for the provision of stochastic signals does not increase in volume as fast as the number of signals, so having established the capability for a large number of signals, the associated arithmetic hardware can increase rapidly with good utilization. Stochastic systems are not feasible in small sizes as for the regular rate systems, but only become attractive when a large number of arithmetic steps are needed.

The type of functions for which stochastic systems are best suited, are those involving linear combinations of many variables, that is, of the form

$$\sum_n \alpha_n x_n$$

where α_n represents a set of coefficients which may be modified at will. This type of function can be synthesized by a regular array of AND and OR gates. Many applications invoke functions of this kind, and a number have appeared in the literature.

Poppelbaum[19] has applied stochastic techniques to video display problems, showing that translation and rotation of images is possible, as well as conformal mapping, convolution, and integral transformation. The number of gates required for reasonable definition can grow quite astronomically, and a figure of over 10^6 AND gates is mentioned for a display matrix of 32×32 elements, using parallel processing. However some sequential scanning can reduce the number of gates to 10^5 for a display matrix of 200×200, and still enable the display flicker rate to be kept low.

Ribeiro[16] has also mentioned the transformation activities, and adds matrix multiplication as another application.

Gaines[17, 20] has applied stochastic rate systems to the process control environment, where the ready facility of multiplication enables adaptive control and filtering to be carried out. The identification and simulation of complex processes involves large numbers of product, sum and integral operations which occupy a very long time on a general-purpose digital machine because of the serial method of processing. A special-purpose stochastic rate system can perform such operations in parallel, and although each function is carried out slowly the overall processing rate is satisfactory. A bandwidth of 10 Hz and an accuracy of 1% are considered to be suitable for many of these applications while an accuracy of 10% is thought to be satisfactory for situations involving feedback.

Stochastic rate systems have only recently come to the fore, so they are not yet widely applied. The applications mentioned above do seem quite promising, so there may well be wider applications in the future if elements and arrays become commercially available. Gaines suggests that stochastic rate systems open up a new vista in process control, enabling many complex algorithms to be used, which were previously unsuitable because of limited computing power.[20]

7.3. *Sampled Systems*

A well-known sampled pulse-rate system is the ubiquitous frequency-meter/counter. The device operates by accumulating input pulses into a counter for a predetermined time. By taking gating and clock signals from various sources, it is possible to calculate pulse-rate ratios, differences, and other simple functions. Such arrangements are very similar to the regular pulse-rate systems mentioned above, except that the pulse-rate irregularity due to b.r.m. elements and direct addition of rates, is eliminated. A repeatable quantized result is obtained, but it is sampled and is only available at the end of the computing period. In this respect the sampled systems tend perhaps to be slower than continuous rate systems, since the sampling period must always allow for the full-scale quantity.

Sampled pulse-rate systems have found some use in the past because they are quite straightforward, and quite difficult functions can be realized with standard circuit blocks. Schmid[21] describes a sampled pulse-rate system based on a simple computing element, which is capable of a wide range of computation functions. The basic element consists of a counter followed by a comparator, there being several inputs which may be gated to provide a choice of inputs or a convergence of several input signals. The device works on two cycles, one during which a signal is accumulated in the counter, and a second during which the counter is emptied in some specific manner in order to produce an output signal. Signals passed between units are of the pulse-width variety, having a repetition period equal to the system sampling period, and having a width proportional to the signal quantity. In addition for example, each input pulse gates a constant clock rate into the counter, each individual input channel employing a clock of different phase such that no two pulses coincide at the input of the counter. During the input period, the counter accumulates a count proportional to the sum of the input signals, and a pulse-width output signal is generated during the next system period by driving the counter down to zero with a constant clock rate.

Integration is carried out by two counters driven in parallel by a constant clock signal, such that each counter cycles completely in one system period. The two counters are set initially in phase, and their phase relationship allowed to be disturbed by the input pulse rate. Input pulses are therefore accumulated algebraically in one of the counters forming the integral, and the resulting phase shift between the two counters is a continuous measure of the integral, obtained without destructive read-out. Other similar techniques are used to produce multiplication, division and differentiation. An interesting further development is the generation of functions such as the circular functions, both by direct generation and by a linear segment approximation technique.

Schmid quotes a speed-accuracy product in practice which is equal to the maximum allowable clock rate. For example, with 1 MHz clock rate, and a 1 ms system period, an accuracy of 1 part in 10^3 should be attainable. However, Ribeiro[22] points out that while an adder/subtractor unit has an output sampling period of T (maximum output giving pulse-width $T/2$), a differentiator has an output period of $3T$, and a multiplier has an output period of $2T$. These factors should be taken into account when assessing the effective bandwidth of the system.

The advantage of this system is that results are more repeatable than with the continuous rate system, for although a b.r.m. is used in certain cases, it is reset to zero at the commencement of a cycle. The application mentioned by Schmid is the conversion of rectangular to polar co-ordinates in navigation systems.

Ribeiro[22] comments on Schmid's system, and gives an improved set of elements which, although operating on a similar principle, does not require conversion to pulse-width signals between units and makes use of more elegant bidirectional counters which can accept simultaneous add and subtract pulse-train inputs. This improvement has the effect of completing every operation in the same time, and is a more satisfactory method altogether.

Gaines and Joyce[23] describe the concept of phase computers which operate on a similar principle to that used by Schmid in the integrator. Unidirectional binary counters are used as elements and, when connected in groups with appropriate gates, can be arranged to transfer data and perform the usual arithmetical operations. The concept is extended well beyond that envisaged by Schmid however, since the problems of programming and sequencing are also dealt with and realized by a universal sequencing unit. The flexibility inherent in this approach means that certain operations may be carried out sequentially, under stored-program control, and so a good balance is struck between wholly parallel and wholly serial processing.

For a clock of 4 MHz, the time for any incremental calculation is given as 250 µs for 10-bit precision, which is adequate for the navigational and process control applications which are envisaged.

7.4. *Digital Differential Analyser*

The digital differential analyser (d.d.a.) is an incremental computing device which calculates the integral of a variable with respect to another variable other than time. The heart of each unit is an accumulator functioning as a pulse-rate integrator, and signals between units are pulse-rates proportional to signal rates-of-change. Figure 18 illustrates the general principle of the d.d.a. integrator. The input rate $\mathrm{d}Y$ is a rate proportional to the rate-of-change of variable Y, and these increments are totalled in an accumulator to form the latest value of Y. Input rate $\mathrm{d}X$ is of similar form to $\mathrm{d}Y$, and at each pulse causes the contents of the counter to be added into register R, i.e. register R contains $\sum Y\mathrm{d}X$. Register R overflows quite regularly since it is only of the same capacity as the Y accumulator, and the overflow pulses form an incremental measure of the integration $dS = YdX$. Interconnection of the basic integrators can enable other operations such as addition, subtraction and multiplication to be carried out.[24]

In the past, use of the d.d.a. has been restricted. Since it operates at a lower speed than a conventional digital processor, a speed advantage may only be achieved by having a number of integrators operating in parallel. Thus a large number of high capacity registers are required, which has tended to be a disadvantage of the d.d.a. Certain processors have been built using parallel tracks on a magnetic drum as storage registers,[24] and have operated well although strictly special purpose and very inflexible. The present interest in l.s.i. is tending to change this outlook, since one of the first l.s.i. packages offered by manufacturers has been d.d.a. package, which may now be regarded as a suitable building block for any system. Interest has also been aroused in the possibility of making a d.d.a. machine more flexible, and Wood[25]

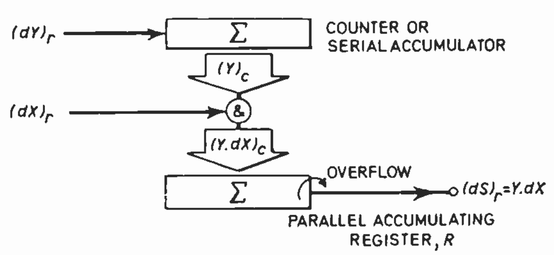


Fig. 18. Principle of digital differential analyser.

describes an approach which will enable a d.d.a. structure to be built with stored interconnexions and hence with great operational flexibility.

Several speakers at the recent IFAC symposium on pulse-rate techniques[12] spoke in favour of the d.d.a. for applications where an analogue computer does not have sufficient accuracy, and a digital computer cannot be justified economically. The trend towards highly parallel processing can be met by the d.d.a. element, particularly in the field of process control.

8. Conclusions

This study of pulse-rate techniques has shown some of the interesting possibilities inherent in this method of signal representation. An attempt has been made to present the unifying principles behind a number of applications which have individually employed several different aspects of pulse-rate working.

The present movement towards the manufacture and use of l.s.i. circuits has introduced requirements for suitable circuit functional blocks having few external interconnexions and many internal gates connected in repeated regular structures. The trend will be towards parallel type structures which will make best use of these complex l.s.i. circuits. Pulse-rate processing circuits are of this general type, and their relatively slow operation may be offset by use of several similar units operating in parallel on different signals. Since pulse-rate signals only occupy one or two signal lines, they are of optimum form for machine structural changes by stored program.

Applications seem to fall into closely defined categories. Regular rate signals are best used in small, special purpose, local processors, where they will offer better accuracy and reliability than the equivalent analogue circuits. They also offer the possibility of programmed or adaptive operation. Stochastic-rate signals seem to be best suited to large special-purpose machines in situations where their low accuracy and speed are not liabilities and where it is necessary to process a large number of signals. Examples are the process control field, and provision of certain visual displays. The sampled systems do not appear to have outstanding characteristics although their use is proposed for the process control and navigational fields. The phase computer is the most attractive in this category. The d.d.a. has a bright future predicted for it, 'particularly in the process control area, where present general-purpose processors are not capable of handling the overall data throughout, in conjunction with complex calculations involving entire systems of differential equations.

It appears that pulse-rate techniques will be more to the fore in the future than in the past, although the extent to which they are adopted will be largely determined by the trend in l.s.i. circuits. The b.r.m. for instance, would be a suitable candidate for l.s.i., but unless it could be used in sufficient numbers it would not be economic. Developments in the field of digital-code arithmetic units are also proceeding fast, and already a parallel multiplier is proposed which could form an operational circuit like the simple feedback circuit of Fig. 14, but with much faster response time.[14]

9. Acknowledgments

The author would like to thank Prof. A. J. Eales, Head of the School of Electrical Engineering, Bath University of Technology, for the use of facilities both in experimental work and preparation of this paper.

Special mention must be made of Mr. J. S. Wright, now of Southampton University, who conducted a searching investigation into the realization of regular pulse-rate circuits as a final year undergraduate project during the academic year 1966–7. The concept of the b.r.d. as an operational element is due to him, as well as the initial concepts of the feedback systems.

10. References

10.1. *Regular-rate Systems*

1. Rey, T. J., 'Digital rate synthesis for frequency measurement and control', *Proc. Inst. Radio Engrs*, **47**, pp. 2106–12, December 1959.
2. Wood, P., 'Frequency meter with continuous digital presentation', *J. Brit. Instn Radio Engrs*, **26**, No. 2, pp. 109–13, August 1963.
3. Phillips, R., 'Many digital functions can be generated with a rate multiplier', *Electronic Design*, **16**, No. 3, p. 82, 1968.
4. Martin, J. D., 'Pulse-rate to digital-code tracking converters', Proc. Conf. on Digital Methods of Measurement, Canterbury, July 1969, pp. 95–104. (I.E.R.E. Conference Proceedings No. 15.)
5. Lundh, Y., 'Digital techniques for small computations', *J. Brit. Instn Radio Engrs*, **9**, No. 1, pp. 439–49, January 1959.
6. Grabbe, E. M. *et al.* (editors), 'Handbook of Automation, Computation, and Control', Vol. 2, ch. 29. John Wiley & Sons Inc., New York. 1959.
7. Yang, H. Z., 'Determination of maximum error of a binary multiplier', *Automation and Remote Control*, **21**, No. 7, pp. 709–13, 1961.
8. Butaev, G. M. and Romashkan, V. S., 'Linear code to pulse-frequency converter', *Measurement Techniques*, No. 2, pp. 194–6, February 1968.
9. Vincent, C. H. and Rowles, J. B., 'A digital linear rate-meter', *Nuclear Instruments and Methods*, **22**, pp. 201–20, 1963.
10. Jager, F. de, 'Delta modulation, a method of p.c.m. transmission using the 1-unit code', *Philips Res. Rep.*, **7**, pp. 442–66, December 1952.
11. Inose, H. *et al.*, 'New modulation technique simplifies circuits', *Electronics*, **36**, pp. 52–5, 25th January 1963.
12. Edwards, R., 'Pulse-rate and pulse-number signals in automatic control', *Control*, **12**, pp. 627–30, July 1968 (This is a report of the IFAC Symposium on Pulse-rate and Pulse-number Signals in Automatic Control, Budapest 1968.)
13. Evers, A. F., 'A versatile digital frequency synthesizer for use in mobile radio communication sets', *Electronic Engineering*, **38**, pp. 296–303, May 1966.
14. Martin, J. D., 'Simulation of function with variable poles', *Electronics Letters*, **4**, No. 26, pp. 594–5, 27th December 1968.
15. Agar, J., 'Frequency modulating transducers', *The Radio and Electronic Engineer*, **38**, No. 2, pp. 89–97, August 1969.

10.2. *Stochastic-rate Systems*

16. Ribeiro, S. T., 'Random-pulse machines', *I.E.E.E. Trans. on Electronic Computers*, **EC-16**, No. 3, pp. 261–76, June 1967.
17. Gaines, B. R., 'Stochastic computing', *Amer. Fed. Inform. Processing Socs, Proc.*, **30**, pp. 147–56. (Spring JCC 1967.)
18. Taylor, J. M., 'Generation of multiple pseudorandom point processes from m-sequences', *Electronics Letters*, **5**, No. 2, pp. 28–9, 23rd January 1969.
19. Popplebaum, W. J. *et al.*, 'Stochastic computing elements and systems', *Amer. Fed. Inform. Processing Socs, Proc.*, **31**, pp. 635–44. (Fall JCC 1967.)
20. Gaines, B. R., 'Techniques of identification with the stochastic computer', Proc. IFAC Symposium on Problems of Identification, 1967.

10.3. *Sampled Systems*

21. Schmid, H., 'Operational hybrid computing system provides analog-type computation with digital elements', *I.E.E.E. Trans.*, **EC-12**, pp. 715–32, 1963.
22. Ribeiro, S. T., 'Comments on pulsed-data hybrid computers', *I.E.E.E. Trans.*, **EC-13**, pp. 640–2, 1964.
23. Gaines, B. R. and Joyce, P. L., 'Phase computers', Proc. 5th A.I.C.A. Congress, pp. 48–57, 1967.

10.4. *Digital Differential Analyser*

24. Mayorov, F. V., 'Electronic Digital Integrating Computers—Digital Differential Analysers' (Iliffe, London, 1964).
25. Wood, P. E., 'Digital differential analysers with arbitrary stored interconnections', *I.E.E.E. Trans.*, **EC-14**, pp. 936–41, 1965.

Manuscript first received by the Institution on 1st November 1968 and in revised form on 9th September 1969. (Paper No. 1291/CC59.)

The Author

J. D. Martin received his B.Sc. (Eng.) degree in electrical engineering from Northampton Engineering College, London (now the City University) in 1957. He subsequently received a M.Sc.(Eng.) degree for postgraduate research on transistor blocking oscillators. His early training was with Kelvin-Hughes Ltd., and from 1960 to 1965 he was concerned with remote control systems for railways and public utilities with Westinghouse Brake and Signal Co., Ltd., Chippenham. In 1965 Mr. Martin was appointed a lecturer in the School of Electrical Engineering at Bath University of Technology.

U.D.C. 621.317.727.2:681.14

A High Speed, High Accuracy, Digitally-set Potentiometer

By

S. MARJANOVIC, Ph.D.†

AND

D. R. NOAKS, C.Eng., M.I.E.E.‡

Summary: A digitally-set potentiometer (d.s.p.) is required for any hybrid computing arrangement where iterative routines are used to establish the solution to dynamic systems. If computing accuracies of 0·1% at signal frequencies of 10 kHz are to be obtained special techniques in the design of a d.s.p. are necessary.

A complete design for a 10-bit d.s.p. to fit this specification is presented, based on the use of a ladder network of low propagation delay. This in turn requires careful design of bipolar switches and a new form of switch is presented which uses f.e.t.s in the feed-back loop of monolithic d.c. amplifiers. The overall behaviour of the d.s.p. is illustrated and an estimation of errors made.

1. Introduction

All hybrid computing techniques require the setting of coefficient potentiometers at the beginning of, and possibly during, an iteration of a solution. The speed of setting is only one of the parameters limiting the iteration rate and a transition from passive contact elements to the use of solid-state devices has naturally occurred as operational amplifier bandwidths have moved into the megahertz region, thus requiring faster digital control of associated devices such as track/store elements, coefficient potentiometers, digital-to-analogue converters, etc.

The design of a digitally-set potentiometer (d.s.p.) is dependent on the two main parameters, maximum demanded iteration rate and dynamic accuracy. The values adopted for the design presented are 10 kHz and 0·1% respectively.

The function of a d.s.p. is to operate on the incoming analogue voltage, v_a, by a digital voltage (i.e. a digital number), v_d, so as to form an output voltage, v_o, given by

$$v_o = k v_a v_d$$

where k is a constant.

To achieve the bandwidth and accuracy adopted above, careful attention must be paid both to the binary weighting configuration chosen and to the design of the digitally-controlled bipolar switches necessary to effect the d.s.p. characterizing equation and a detailed discussion of these now follows.

2. Principles of Operation and Accuracy

The digital-to-analogue converter (d.a.c.) is a special case of a d.s.p. since by keeping the analogue input voltage constant the output voltage becomes proportional only to the digital signal. Thus the resistive network configurations normally used in a d.a.c. can also be employed in a d.s.p., and a ladder configuration of the type shown in Fig. 1 is chosen because of its superior performance with respect to propagation delay.

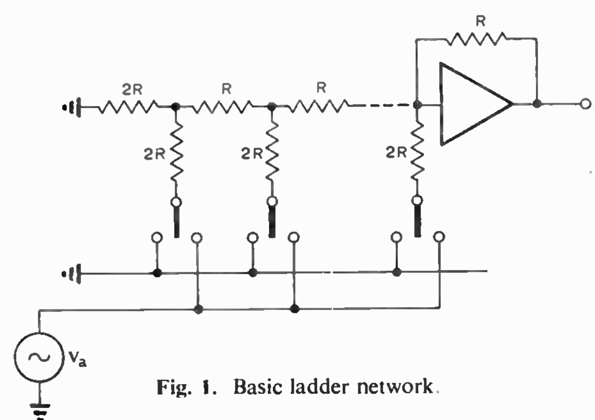


Fig. 1. Basic ladder network.

The static limitations of such networks are well documented[1] and only the dynamic effect of propagation delays along the ladder will be considered. This delay is caused by the presence of reactive components associated not only with the manufacturing of the resistors, but also with their mounting on the board which introduces stray capacitances. The magnitude of this delay determines the upper frequency of the analogue signal which can be handled with a given accuracy.

If a uniform delay of Δt seconds per section of the ladder is assumed, the output voltage V_o for a sinusoidal input voltage is

$$V_c = \frac{1}{2} V_a \cos \omega t + \frac{1}{4} V_a \cos \omega (t+\Delta t) + \ldots + \frac{1}{2^n} \cos \omega\{t+(n-1)\Delta t\}$$

† Formerly at the University of Birmingham; now with the Institute for Automation and Telecommunications, Belgrade, Yugoslavia.

‡ Department of Electronic and Electrical Engineering, University of Birmingham.

where V_a is the amplitude of the analogue input voltage.

If the error voltage is ΔV then the maximum percentage error is given by

$$\frac{\Delta V}{V_o} = \frac{1}{2^2}\,\omega\Delta t + \frac{1}{2^3}\,2\omega\Delta t + \ldots + \frac{1}{2^n}\,(n-1)\omega\Delta t$$

$$\simeq \omega\Delta t$$

For an accuracy of 0·1% up to signal frequencies of 10 kHz, the maximum delay as calculated from the last equation is 16 ns per section. With an experimental model of a ladder network using 5 kΩ metal film resistors, a delay of $\Delta t = 40$ ns per section was obtained. It is reasonable to expect that within the small range of values of the ladder resistor the delay stays proportional to these values and that a ladder using 2 kΩ resistors is an appropriate choice to handle 10 kHz signals.

The propagation delay with respect to the digital input has little significance because it is overshadowed by the duration of the spikes, caused by the state changing of the switches, which last much longer.

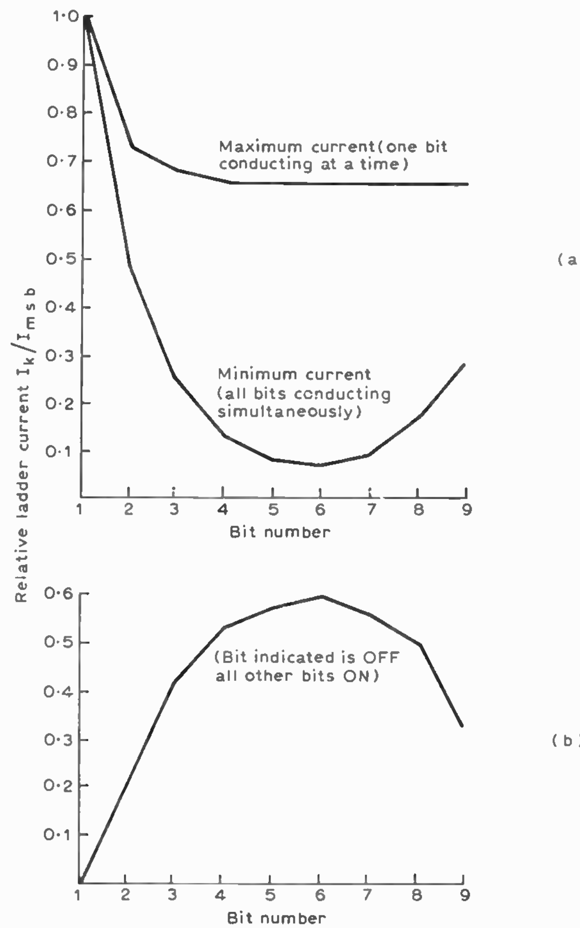
3. Requirements of the Ladder Network Switch

In order to preserve the binary relationship between the currents in individual branches of the ladder it is necessary that when the digital setting of a switch is zero, or OFF, the branch concerned is returned to some reference voltage (normally ground). Therefore, the types of switch required for application with the ladder network must be double throw with equal resistance when connected to either the analogue input or to the reference potential of the ladder.

A knowledge of the following design parameters is thus necessary:

1. The maximum resistance of the switch when conducting.
2. Maximum forward current through the switch between the ladder and the input voltage.
3. Maximum current in the reverse direction through the conducting switch which provides the path to the ground.
4. Maximum reverse voltage.

It is necessary to know the value of the switch resistance when a ladder resistor is connected either to the input voltage or to the ground potential as this enables trimming of the ladder resistors for the purpose of achieving the necessary tolerance for the required accuracy of operation. For a ladder network with $R = 2$ kΩ and assuming all other errors to be negligible, a maximum variation of the switch resistance of 2 Ω is permissible for an overall accuracy of 0·1%.



(a) Maximum and minimum source currents of switches when ON.

(b) Sinking current of switches when OFF.

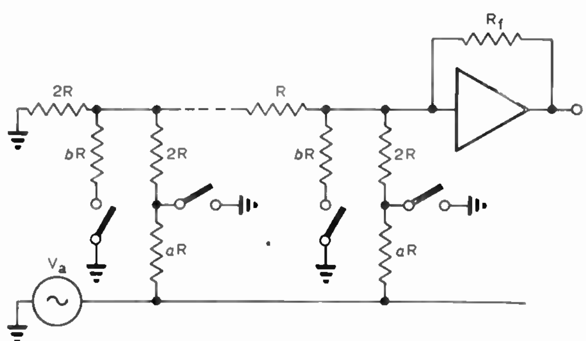
Fig. 2. Distribution of current in ladder network.

When designing a switch, estimates must be made in assessing the maximum current which flows through it when the corresponding ladder resistor is connected either to the input voltage or to ground. The current is a function of the states of the ladder switches, and the distribution of current through a 10-bit ladder network is shown in Figs. 2(a) and (b)[2] where the bit current I_k is plotted in normalized form as I_k/I_{msb} against the bit number. (I_{msb} is the most significant bit current.) In the case of Fig. 2(a) the direction of this current is from source to ladder and in Fig. 2(b) it is the reverse current which flows into an OFF switch when all other bits are ON. It can be seen from these current distributions that the magnitude of the maximum currents through the switch when it connects the ladder with the input and when it connects it with the ground are approximately the same, amounting to about 60% of I_{msb}.

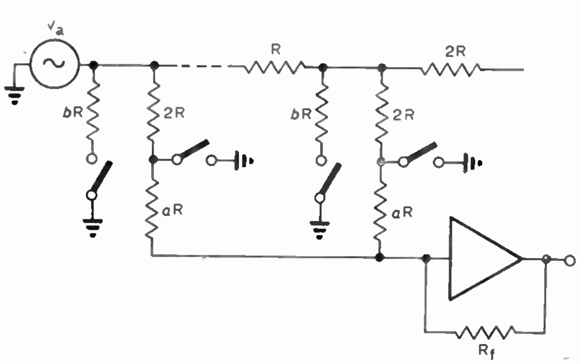
4. Modified Ladder Network

The switch which is required for the ladder network of Fig. 1 has to be of the series-shunt type. It is not difficult to see that the specification of the conditions which the switch has to meet is very extensive and that some of these conditions cannot be satisfied with existing switching elements available on the market; this is particularly true with respect to a very low value of the ON resistance, which is only available in a m.o.s.t. at a considerably increased cost.

The design of a shunt-type switch presents a less serious problem, a very low ON resistance when switched to ground can be realized although this requires a modification of the structure of the ladder network to accommodate it, the simplicity of the existing ladder then being sacrificed in exchange for reduced switch versatility. A compromise has to be made as the cost of high precision resistors tends to be very high, but it seems that there is no alternative at the moment.



(a) Constant output impedance.



(b) Constant input impedance.

Fig. 3. Practical forms of ladder network.

Two types of modified ladder network are shown in Fig. 3(a) and (b). Both use twice as many precision resistors per section as that of the basic ladder network. This is because one of the resistors of each section of the original ladder is split into two branches formed of three resistors which are alternatively switched ON and OFF. The number of combinations of values which these three resistors can have while preserving the value of $2R$ is clearly unlimited, but a choice has to be made to suit the switch requirements. For equivalence the relation which the resistors must satisfy is given by

$$b = (4+2a)/a$$

The values for a and b can be chosen on the bases of criteria which involve the loading of the analogue source, the uniformity of the resistance values of the ladder and finally error.

Consider these effects in relation to the network of Fig. 3(a). The worst case loading of the analogue source is when all resistors connected to the analogue input are grounded, i.e. when the digital signal comprises all zeros. The maximum current which the analogue source must supply is then given by

$$\frac{nV_a}{aR}$$

where n is the number of bits of the ladder.

This current is always larger than the current which has to be supplied to the ladder when the digital signal comprises all ones. For the modified ladder network the bit current, i_j, i.e. the current which flows from the source to each ladder section, is related to the bit current, i_k of the original ladder by the expression

$$\sum_{j=1}^{n} i_j = 2\sum_{k=1}^{n}\left(\frac{I_{\text{msb}}}{b} + \frac{1}{a+2}\, i_k\right)$$

where $I_{\text{msb}} = V_a/(a+2)R$.

The worst current demand of the ladder can now be found by reference to Fig. 2(a).

The offset voltage at the output of any switch when ON supplies a current into the ladder which introduces an offset current given by

$$\sum_{k=1}^{n} \frac{V_{\text{offset}\,k}}{2^{k-1}R}$$

where $V_{\text{offset}\,k}$ is the offset voltage of each switch. It is important, however, that the offset current of any bit does not exceed the value of one-half of the least significant bit current, namely $I_{\text{msb}}/2000$ (assuming all other errors to be negligible) otherwise the linearity of the binary relationship will be impaired; for the same reason the OFF resistance of any switch must not be less than $2000R$.

The maximum impedance of any switch when conducting can be determined from the maximum permissible offset voltage, and the maximum current which the switch has to sink. The current flowing into the switch is equal to the sum of the current from the switch output and the current which comes from the analogue source via the ladder network. For worst case design the maximum current from the ladder can be found with the help of Fig. 2(b) which shows the current which flows into any OFF switch when all of the other switches are ON. In a 10-bit ladder network, the worst case occurs at the sixth bit (in a symmetrical ladder network it is at the bit in the middle of the network) and the maximum current the switch sinks is

$$i_{max} = V_a/aR + 0{\cdot}6V_a/(a+2)R$$

The maximum value of the switch impedance is therefore,

$$|Z_o| = \frac{V_{offset}}{V_a/aR + 0{\cdot}6V_a/(a+2)R}$$

As seen from the output of the analogue source the ladder represents a varying load which may become very troublesome particularly when the value of a has to be chosen to be relatively low. This imposes rather stringent conditions on the magnitude of the output resistance of the analogue source over the whole frequency range of interest, giving a maximum permissible value for 0·1% accuracy of

$$\frac{10\times 10^{-3}}{2\sum_{k=1}^{n}\left(\frac{I_{msb}}{b}+\frac{i_k}{a+2}\right)-n\frac{V_a}{aR}}$$

The configuration of the ladder network of Fig. 3(b) is identical to the previously described network, except for the input and output connections which have their positions interchanged. As a result of this interchange two major differences ensue:

1. The input impedance of the second version of the modified ladder network becomes constant at the expense of the output impedance, which was originally independent of the position of the switches; this helps in relaxing the severity of the requirements relating to the output impedance of the analogue source.
2. In the first version of the ladder the total offset current due to the existence of offset voltages, assuming that the offset voltages are of the same magnitude, is V_{offset}/R, while in the second version it is $nV_{offset}/2R$.

5. The Shunt Switch Design

A switch configuration has been devised which exploits low-cost m.o.s.t.s used in conjunction with low-cost high-gain operational amplifiers. The switch has been designed for the modified ladder networks of Figs. 3(a) and (b) allowing operation up to 10 kHz with 0·1% accuracy.

Metal oxide transistors exhibit excellent characteristics with respect to many switching parameters including drift-free operation, negligibly small current between the gate and channel, very large OFF resistance and low gate-drain capacitance. However, their ON resistance is usually not smaller than two or three hundred ohms which is, for many switching applications, unacceptable. The direct use of a m.o.s.t. as a switch in the ladder network would require large resistor values which then will limit severely the frequency range of operation.

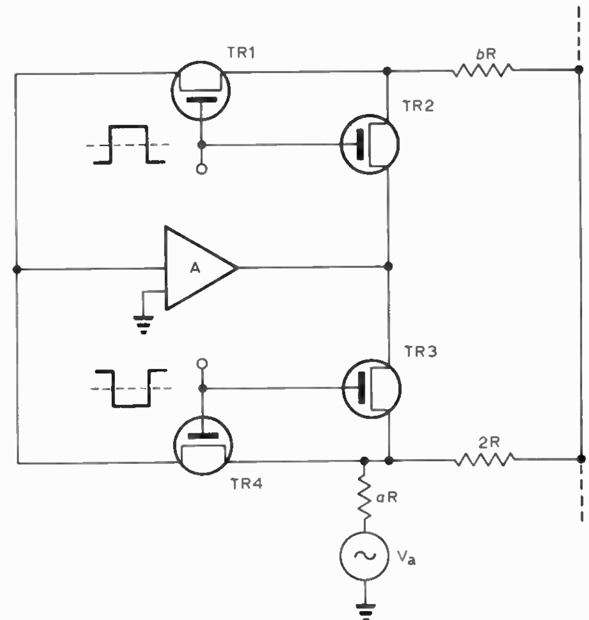


Fig. 4. Principles of bipolar shunt switch.

The switch of Fig. 4 incorporates transistors in the feedback loop of the monolithic amplifier A, thus reducing the value of the ON resistance by the amplifier loop gain. While transistors TR1 and TR2 conduct simultaneously transistors TR3 and TR4 are switched OFF. The switch thus provides two points which are alternately either open circuit or at ground potential and only one such switch per bit is needed for the modified versions of the ladder. Except for the short intervals during the transients, the amplifier operates in a linear mode which reduces the time interval when the amplifier is saturated to the value determined by the spike widths produced by the switching transistors.

Since the output impedance of the amplifier depends on its gain bandwidth characteristics, it is clear that the output voltage will take some time to reach the steady-state value.

The form of the transient response is given in the Appendix and assuming an amplifier with a gain bandwidth product of $s_{10} = 6$ Mrad/s and an output resistance of $R_o = 250\ \Omega$ to be used in a ladder with $R = 2{\cdot}5\ \text{k}\Omega$ and switch transistors of $R_{on} = 250\ \Omega$, the time required for the output voltage of the amplifier to set within $0{\cdot}1\%$ of the input voltage is about 200 ns.

The high-frequency leakage across the non-conductive transistor can usually be neglected with m.o.s.t.s because the interelectrode capacitance of the transistor are comparable with the stray capacitances of the ladder network itself.

Switching spikes appear at the output and are caused by the leakage from the gate into the signal path. It is very difficult to present an exact analysis of these since the resistance of m.o.s.t.s depends on the finite rise-time of a rapidly varying gate voltage. For a qualitative treatment it can be assumed that this resistance is constant when the transistor is conducting and that it is infinitely large when it is OFF. Consider the case when transistors TR1 and TR2 of Fig. 4 are being switched from the ON into the OFF state, by the gate voltage of amplitude V. A negative charge of $2VC_{dg}$ coulombs is transferred to the output and is partially absorbed by the combination of C_{sd1} and the input capacitance of the amplifier. During short transients the amplifier can be regarded as a capacitor at the input and an inductor at the output. The drain-source capacitance of TR2 and equivalent output inductance form a series tuned circuit with a resonant frequency of several hundred megahertz. This remaining charge at the output of TR1 and TR2 decays with a time-constant determined by the equivalent capacitance and equivalent resistance at this point. The height of the spike is roughly given by the ratio of the junction capacitances between the drain and gate and between the drain and source.

In practice the area under the spikes of the output voltage waveform allows determination of the charge per switching cycle produced in the ladder and in the worst case this is equal to twice the charge of an individual bit. On the basis of the integrated charge per switching cycle it is possible to calculate the maximum speed of setting of the digital potentiometer beyond which the switching spike contribution becomes excessive. The maximum rate of switching is then given as

$$f_{max} = \frac{1}{\Delta T} = \frac{10^{-3} I_{msb}}{2Q}$$

where

$$Q = \frac{\text{area under the spike of the individual switch}}{2R}$$

The presence of the switching spikes is not critical when the digitally-set potentiometer is to be used only once per computing run as in the case when the gain of the amplifier is to be altered; however, when used as an analogue multiplier, the speed of setting is of extreme importance.

To minimize the switching spikes it is necessary to build the whole network on a board with a ground plane as well as to use the shield between the gate and drain of the switching transistor to reduce the charge which otherwise may be induced.

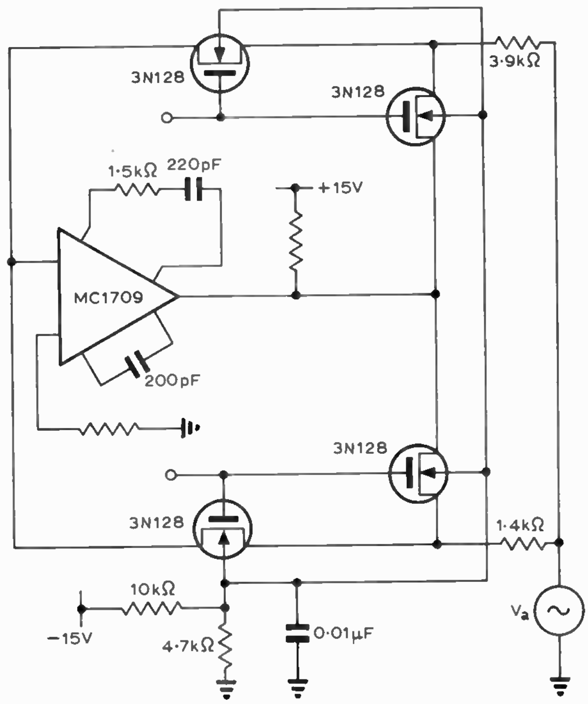
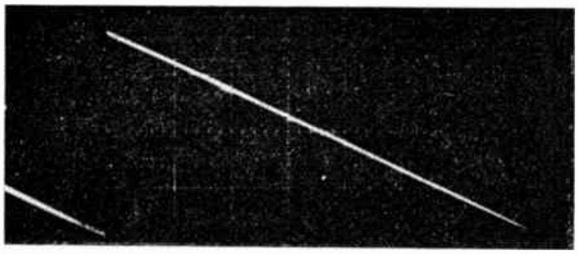


Fig. 5. Practical design of shunt switch.

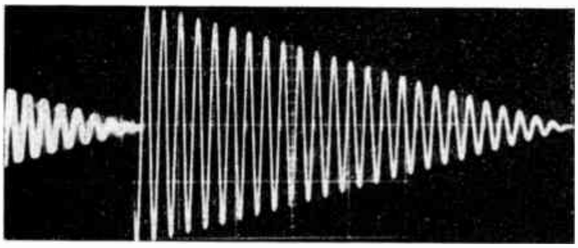
6. Practical Realization of the Digitally Set Potentiometer

Figure 5 shows the diagram of the switch which has been built. The interface circuit, designed to be driven by integrated logic, has been added to provide voltage levels compatible with the switching requirements of the m.o.s.t.s. The transistors are RCA 3N128 with a typical drain-gate capacitance of 0·2 pF.

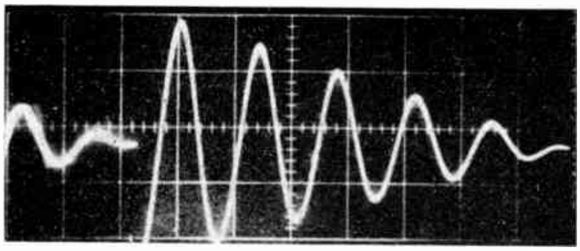
The switch is made to operate for signal levels of ± 5 V, as full input voltage is never applied to any switch. The current-sinking capacity depends on the output compensation of the amplifier used and in this instance was ± 4 mA. Reduction of the compensating capacitor, as much as it is beneficial for the sinking of current, is not advisable since it unmasks the high frequency poles of the gain function which then affects the nature of the amplifier output impedance. Also the required gain at the highest signal frequency must be known to establish correct values of compensating elements to ensure that the output impedance is kept below the level which can be tolerated. The amplifier used is Motorola type MC1709CP.



(a) Input voltage 10 V d.c.
Vertical scale 3 V/cm.
Horizontal scale 200 μs/cm.



(b) $f = 20$ kHz.
Input voltage 10 V sinusoidal.
Vertical scale 5 V/cm.
Horizontal scale 200 μs/cm.



(c) $f = 300$ Hz.
Input voltage 10 V sinusoidal.
Vertical scale 5 V/cm.
Horizontal scale 2·5 ms/cm.

Fig. 6. Waveforms in digitally-set potentiometer.

For the ladder network of Fig. 3(a), chosen because of better performance with respect to noise, the voltage waveforms at the output of the digitally-set potentiometer with the digital input supplied by a step counter are shown in Fig. 6 for (a) 10 V d.c. input voltage, (b) and (c) 10 V sinusoidal voltage of two different frequencies at the input. The values of $a = 4$ and $b = 3$ were selected as a compromise between the noise which can be tolerated and the signal power demand from the analogue signal source.

The following specification describes the performance of the digitally-set potentiometer:

Number of bits of the ladder network 10.

Accuracy of the setting 0·1% ± ½ l.s.b.

Estimated maximum rate of setting 30 000 per second for an accuracy of 0·1%.

Estimated accuracy with respect to the analogue input 0·1% up to 10 kHz.

Maximum analogue input voltage ± 10 V.

Maximum current demand from the analogue source 12·5 mA.

Digital voltage levels compatible with either r.t.l. or d.t.l. integrated logic when driven from gates whose outputs are solely loaded by the d.s.p.

Power supplies + 15 V, − 15 V together with the power supply required for the logic.

Power consumption approximately 5 W.

7. Conclusions

In many hybrid computer applications, such as parameter optimization, fast and accurate analogue multiplication and function generation, there is a need for a fast and accurate digitally-set (controlled) potentiometer. The design of such a potentiometer has been presented and its performance evaluated. An accuracy of 0·1% for analogue signals from d.c. up to 10 kHz has been achieved with a maximum rate of 30 000 settings per second.

The resistive ladder network, most frequently used with digital/analogue converters has been modified so as to avoid the use of series switches which are difficult to achieve with high accuracy. A shunt switch has been designed which exhibits a low ON resistance of less than 1 Ω for signals up to 10 kHz. The switch utilizes a switched high-gain monolithic amplifier with m.o.s.t.s as switches placed in the feedback loop.

The design of the whole digitally-set potentiometer does not involve any manual adjustments if a static accuracy not better than 0·1% is required. Constant reduction in the prices of the active components (transistors and monolithic amplifiers) which are used to build this kind of the potentiometer would make high precision resistors the only cost loading components in the entire instrument.

8. Acknowledgment

The authors wish to acknowledge that the work was carried out in the Department of Electronic and Electrical Engineering of the University of Birmingham and forms part of the programme in hybrid computing techniques currently receiving support from the Science Research Council.

9. References

1. Dutta Roy, S. C., 'Terminal impedances and transfer functions of ladder networks', *Proc. Instn Elect. Engrs*, **111**, pp. 1653–1658, October 1964. (I.E.E. Paper No. 4546E, 1964.)
2. Pearman, C. R. and Popodi, A. E., 'How to design high-speed d/a converters', *Electronics*, **37**, No. 8, pp. 28–32, 21st February 1964.

10. Appendix

The output impedance of the amplifier at the points where the ladder is connected is given by

$$Z_{\mathrm{ON}} = \frac{R_d + R_o}{1 + G(s)H(s)}$$

where $G(s)H(s)$ is the loop gain,

R_d is the ON resistance of the switching transistor,

and R_o is the d.c. open-loop output resistance of the amplifier.

The feedback ratio $H(s)$ equals unity because there is no resistor in the input. If a single pole in the frequency response of the amplifier is assumed, the output impedance becomes

$$Z_{\mathrm{ON}} = (R_d + R_o)\left(\frac{s_o}{s+s_{10}} + \frac{s}{s+s_{10}}\right)$$

where s_o denotes the pole
and s_{10} the gain-bandwidth product of the amplifier.
The time response to the step input via the ladder resistor R can be found as the inverse Laplace transform of

$$V_{\mathrm{ON}} = \frac{(s+s_o)(R_d+R_o)}{s(R+R_d+R_o)+s_{10}R}$$

and gives

$$\frac{V_{\mathrm{ON}}(t)}{V_{\mathrm{IN}}} = \left\{\frac{R_d+R_o}{A_o R} + \frac{R_d+R_o}{R+R_d+R_o} - \frac{R_d+R_o}{A_o R}\right\} e^{-t/T}$$

$$\cong \left(\frac{R_d+R_o}{A_o R} + \frac{R_d+R_o}{R}\right) e^{-t/T}$$

where the time-constant

$$T = \frac{R_d+R_o+R}{R}\,\frac{1}{s_{10}}$$

$$\cong \frac{1}{s_{10}}$$

Manuscript first received by the Institution on 30th April 1969 and in final form on 9th September 1969. (Paper No. 1292/IC17.)

Conferences of Engineers

Second General Assembly of W.F.E.O.

107 representatives of the engineering profession from 53 nations and 5 regional federations of engineering societies met in Unesco House, Paris, from 28th to 30th October, 1969, at the Second General Assembly of the World Federation of Engineering Organizations. The Federation was founded and held its First General Assembly in March, 1968.

Representative engineering societies of the following countries were elected to membership of the Federation:

Algeria	Columbia	Indonesia
Libya	Panama	Peru
Puerto Rico	Salvador	Union of Soviet Socialist Republics

The following officers were appointed:

President:

Dr. Eric Choisy (Switzerland) (Re-elected)

Vice-Presidents:

Mr. Robert Gibrat (France) (Re-elected)

Academician A. Samarin (Union of Soviet Socialist Republics)

Mr. C. R. Végh Garzón (Pan-American Union of Associations of Engineers)

Secretary-General:

Dr. G. F. Gainsborough (Commonwealth Engineering Conference) (Re-elected)

Other members of the Executive Committee were appointed as follows:

Mr. A. Gajkowicz (Poland) (Re-elected)
Mr. B. P. Kapadia (India)
Professor V. Péevsky (Bulgaria) (Re-elected)
Professor J-C. Piguet (European Federation of National Engineering Associations)
Mr. M. Sakr (Federation of Arab Engineers) (Re-elected)
Dr. W. H. Wisely (United States of America) (Re-elected)

The United Kingdom representation consisted of the Chairman and Secretary of The Council of Engineering Institutions (Sir Leonard Drucquer and Mr. M. W. Leonard respectively), and in addition Drs. D. F. Galloway (I.Mech.E.), J. H. Jellett (I.C.E.), and D. H. Sharp (I.Chem.E.), Professor J. M. Meek (I.E.E.) and Mr. K. H. Platt (I.Mech.E.) and Mr. J. G. Watson (I.C.E.); additionally Sir Leonard Drucquer and Dr. G. F. Gainsborough (I.E.E. and Secretary-General W.F.E.O.) represented the Commonwealth Engineering Conference.

The General Assembly was gratified to note that since its previous meeting, the Federation had been granted Consultative Status B by UNESCO and Consultative Status by UNIDO.

After discussing the future role of the Federation, the General Assembly agreed that, while encouraging the formation and supporting the activities of regional federations of engineering societies, the Federation should undertake projects in support of the programmes of intergovernmental world organisations and provide a world forum for the discussion of matters of concern to the engineering profession and to societies of engineers.

In the course of an address of welcome to the delegates, given on behalf of the Director-General of UNESCO, Mr. A. Evstafiev outlined studies in the fields of education and training of engineers, and of engineering science and research, which UNESCO hoped that the Federation would be willing to undertake in collaboration with UNESCO. The General Assembly readily agreed to collaborate in this way.

The General Assembly also agreed to undertake on behalf of UNIDO, a study of the role of engineering societies in industrial development; and to organise, with the assistance of UNESCO, an international meeting of editors of journals of engineering societies to discuss matters of common interest, including the possibility of a measure of standardization of practices in areas in which this might be an advantage.

Decisions were also taken to entrust the Executive Committee with the setting up of standing committees to deal with the following matters:

(a) A committee on Education and Training to promote and manage studies in this field, in particular those to be undertaken in collaboration with UNESCO, to which reference was made in Mr. Evstafiev's address;

(b) A Committee on Engineering Information to identify the deficiencies in existing facilities for the dissemination and retrieval of technological information required by engineers and to promote means of remedying them;

(c) A Committee on Ocean Engineering to promote international study in this field and, in particular, to collaborate with other national and international bodies in the International Decade of Ocean Exploration which is to be launched in 1971 following an initiative by the United Nations;

(d) A Committee on Environmental Engineering to promote the study, at an international level, of problems such as the pollution of the air, sea and rivers.

A code of professional conduct for engineers was agreed, and was recommended as a model and general guide in the framing of codes of conduct used by National and International Members of the Federation.

It was agreed that the Third General Assembly of the Federation should be held in Varna, Bulgaria in June 1971, at the invitation of the Bulgarian Scientific and Technical Union.

Seventh Commonwealth Engineering Conference

The Seventh Meeting of the Commonwealth Engineering Conference was held in New Delhi from 2nd to 7th November 1969 at the invitation of the Institution of Engineers, India, under the Chairmanship of the President of the Institution, Mr. T. R. Gupta.

The Commonwealth Engineering Conference was set up in 1946 on the initiative of the Institutions of Civil Engineers, Mechanical Engineers and Electrical Engineers in the United Kingdom. Its meetings, attended by the Presidents and Secretaries of the participating Institutions, are held at approximately 4-yearly intervals; its function is advisory, and it provides a forum in which the participating Institutions can discuss their common problems. Its previous meetings have been held in London, 1946; South Africa, 1950; London, 1954; Australia and New Zealand, 1958; Canada, 1962; and London, 1966.

Sir Leonard Drucquer (Chairman) and Mr. M. W. Leonard (Secretary) represented the Council of Engineering Institutions; the United Kingdom representation additionally consisted of Drs. G. F. Gainsborough (I.E.E.), D. F. Galloway (I.Mech.E.), J. H. Jellett (I.C.E.), Professor J. M. Meek (I.E.E.), Mr. K. H. Platt (I.Mech.E.) and Mr. J. G. Watson (I.C.E.).

The Inaugural Session was held on November 3rd, when the Conference was honoured by the presence of the President of India, Shri V. V. Giri, who gave the inaugural address.

Over 300,000 members of all the professional engineering institutions in the Commonwealth covering various disciplines were represented at this Conference, and Mr. Giri called on scientists and engineers to play a bigger role and to bear a greater burden in removing the suffering of humanity. Technologists and engineers, Mr. Giri said, have the responsibility of applying science to the good of man. He made a plea for giving planning and technology their due place in national development. Eradication of poverty and raising the standard of living can only be achieved, he said, by translating the technical planning and technological advancements into something of practical utility in a way so that they have a wider acceptance specially in rural areas. Mr. Giri urged for the development of agro-industries on a mass scale to solve the problem of poverty and unemployment.

The Conference was attended by representatives of the engineering institutions from Australia, Canada, Ceylon, East Africa, India, Jamaica, New Zealand, Pakistan, Singapore, Trinidad and Tobago and the United Kingdom; the engineering institutions of Ghana, Jamaica and Singapore were elected to the membership of the Conference at the beginning of the first session.

Topics discussed at the business sessions included the qualification of professional engineers and of technical supporting staff, and the statutory registration of engineers.

An important action was to agree to set up a Commonwealth Board of Engineering Education and Training to facilitate mutual assistance between the countries of the Commonwealth in the field of engineering education and training, as for example, by administering funds to enable members of University staffs to visit teaching establishments in other parts of the Commonwealth, to give or to receive information and advice on educational problems and methods. Mr. John Chadwick, C.M.G., Director of the Commonwealth Foundation, attended the session of the Conference at which this matter was discussed, and announced the Foundation's generous decision to give financial support to the Board for a period of three years. The details of the scheme are to be worked out between the Executive Committee of the Commonwealth Engineering Conference and the Officers of the Commonwealth Foundation.

A well-attended one-day open session of the Conference was held on November 4th on the theme 'Priorities in Public Works in Developing Countries', when addresses were given on four aspects of the subject. Dr. K. L. Rao, Minister for Irrigation and Power, Government of India, spoke on financial and related questions. He said that today it was imperative that developing countries should make rapid progress in order to ensure peace, security and stability in the world. He added that it was possible to effect a quick transformation of the developing countries and for this, the concept of a world community has to be fostered. Dr. Rao pointed out that in the developing countries there were conditions of extreme poverty with all the attendant evils. He warned of a reduction in financial aid from developed countries and said that on their part, the developing countries have to realize that development of a country was essentially indigenous and that they should not depend on foreign capital.

Other speakers covered other aspects of the subject: Mr. L. G. Grandy presented a paper by Mr. W. P. Smith of the Montreal Engineering Company Limited on Material Resources; Major-General Harkirat Singh, Past-President of the Institution of Engineers (India), spoke on Manpower Resources, and Mr. Ralph Freeman, Senior Partner, Freeman, Fox & Partners, spoke on Communications. These addresses were followed by brief discussions, and the session was summed up by Prof. J. W. Roderick of the University of Sydney.

On November 5th, the Institution of Engineers (India) held a one-day Seminar on 'The Role and Problems of Professional Institutions in a Developing Economy'. The timing of the Seminar enabled the delegates to the Commonwealth Engineering Conference to attend and to participate in the discussions. The opening address by Mr. V. K. R. V. Rao, Union Minister for Education, Government of India, was followed by brief lectures from Dr. B. D. Nag Choudhuri, a member of the Planning Commission, Government of India and from Dr. K. L. Rao, a Past President of the Institution of Engineers (India).

After the termination of the business sessions, delegates to the Conference visited Calcutta and Bombay, and were generously entertained by the Institution of Engineers (India), by Ministers of the Government of India, and by the Corps of Electrical and Mechanical Engineers. A highlight of the social programme was a visit to Agra to see the Taj Mahal and other historic buildings.

U.D.C. 535.374: 621.375.8

High Peak Power from a CO_2 Laser System

By

R. C. CRAFER, B.Sc.,†

Professor A. F. GIBSON, B.Sc., Ph.D., A.Inst.P.‡

AND

M. F. KIMMITT, B.Sc., A.Inst.P.‡

Reprinted from the Proceedings of the I.E.R.E. Conference on 'Lasers and Opto-Electronics' held at the University of Southampton on 25th to 28th March 1969.

Summary: A CO_2 laser oscillator/amplifier system which has produced peak output powers of 350 kW in 0·4 μs long pulses is described. The output power achieved is shown to be in agreement with a theoretical equation. From these results and from the equation a new laser system has been designed, which it is hoped will produce more than 1 MW peak power. This new design is discussed. One problem associated with high gain amplifiers is their tendency to oscillate. An optical component has been designed which largely overcomes this problem. Basically it consists of a moving tape, opaque to 10·6 μm radiation, which is perforated by the rising edge of the Q-switched pulse.

Following some preliminary studies of the output power available at 10·6 μm from Q-switched CO_2 lasers and on the gain of amplifiers, a large oscillator-amplifier system was constructed to study the problems of achieving very high peak power. The availability of a 40 m long optical bench removed any necessity to fold the system or to obtain particularly high gain per unit length. The preliminary experiments had indicated that a.c. driven amplifiers gave rather better gain than d.c. driven tubes, especially for large diameters, and as high voltage a.c. supplies could be made much more cheaply and more quickly than d.c. ones, it was decided to use this form of excitation. The chief problem this caused was the need to synchronize, both in frequency and phase, the rotation of the Q-switching mirror with the mains supply. (There is as yet no other really satisfactory method of Q-switching CO_2 lasers). A limitation of the a.c. system is that one is restricted to a maximum of 100 pulses per second—and in our eventual design to 50.

Considerable care was taken in constructing an oscillator which only oscillated on axial modes and so produced a reasonably well collimated beam. This involved using a low Q system which oscillated only for a short period when the rotating mirror was exactly in line and also meant that oscillation only took place on the one or two rotational lines with the highest gain. The final design was a 3 m long, 2·5 cm diameter oscillator. In order to avoid rotating a curved mirror, or coupling out through a curved semi-reflecting mirror, a rotating plane mirror was set at 45° to the tube axis. This mirror was made of aluminized stainless steel and it reflected the radiation on to a 10 m radius of curvature gold coated glass mirror set at 90° to the tube. The output end of the cavity was a plane disk of Irtran II which has a reflectivity of only 14% and a transmission of rather less than 70%. With 100 Hz rotation speed, this oscillator produced pulses of 0·4 μs duration and 600 W peak power. The pulses were triangular with a time to peak of 0·15 μs and the output beam was 6 mm across. The radiation then passed through 6 m of 2·5 cm diameter amplifier which amplified the pulses to about 25 kW. Further stages of amplification followed, 3 m of 3·75 cm diameter tube and 6 m of 5 cm diameter and the final output was 200 kW, which was later increased to 350 kW. The total length of the oscillator-amplifier was 21 m, including gaps between sections and short 'dead' sections which were necessary to ensure uniform excitation of the tubes.

All the tubes were water-cooled and the windows between the various amplifier sections were of NaCl, which were angled at 35° but not 'Brewsterized'. The a.c. supply to the oscillator and amplifiers was from saturating high voltage transformers and it was found, and later confirmed,[1] that the maximum gain of the 2·5 and 3·75 cm diameter tubes occurred at approximately the same time during the mains half-cycle. The maximum gain of 5 cm tubes was much earlier and a phase shifter was therefore incorporated into the transformers exciting this amplifier section. The breakdown voltage of the tubes, which were excited in 3 m sections, was about 10 kV and the operating voltage 5–6 kV. It proved necessary to water-cool the cathodes and with the a.c. system this would have involved cooling both electrodes and using an electrically insulating coolant. It was therefore decided to insert a rectifier into each power supply to suppress one half-cycle and this allowed the water-cooled cathodes to be grounded. The gas—in the first instance a 'standard' mixture of 0·6 torr CO_2, 0·8 torr N_2 and 4·2 torr He—entered the tubes at the high voltage and was pumped out, with about 30 gas changes per minute, at the cathodes.

† Formerly at the University of Essex; now at the University of Hull.

‡ Department of Physics, University of Essex, Wivenhoe Park, Colchester, Essex.

The immediate problem when the complete system was excited was that the amplifier oscillated. There was sufficient reflectivity between the oscillator end mirror and various surfaces (tube couplers etc.) within the amplifier to provide enough feedback for oscillation. This was not altogether surprising as the small-signal gain of the amplifier system was between 30 and 40 dB. Oscillation was overcome by fitting six stops varying from 5 mm diameter at the entrance of the first amplifier to 4 cm diameter at the end of the final stage. However, unless the laser system was directed at a non-reflecting surface, oscillation still occurred.

A second problem was that with the 'standard' gas mixture an output of only about 240 kW was reached and the gain of the final amplifier section was down to about a quarter of its small signal gain, indicating very severe saturation at about 20 kW cm^2. Increasing the pressure of He to 10–12 torr in the final amplifier section increased the output by about 50% to 350 kW or about 30 kW/cm^2. At this level of saturation one would expect to have reached power levels an order of magnitude higher assuming the c.w. saturation level of Holtz and Austin[2] and an upper level lifetime of 1 ms. It is now clear that the output was limited by the relatively slow exchange time of the rotational states of CO_2 and the saturation level is entirely consistent with our later observations and calculations. Using eqn. (1) of our companion paper[3]:

$$\frac{G}{G_0} = \left(\frac{P}{P+\omega}\right) \exp\{-qt(P+\omega)\} + \left(\frac{\omega}{P+\omega}\right) \exp\left(-\frac{P\omega t}{P+\omega}\right) \qquad \ldots\ldots(1)$$

and assuming a power of 28 kW/cm^2, $q = 20$ (one line from the oscillator), $\omega = 10^6\ s^{-1}$ (approximately correct for 10–12 torr He) and $t = 0{\cdot}15 \times 10^{-6}\ s^{-1}$ (time to peak of pulse which is not strictly the t of eqn. (1)), then $G/G_0 = 0{\cdot}25$ which is near enough the observed result. With 4·2 torr He ω is reduced and again the saturational level of 20 kW/cm^2 is in good agreement with eqn. (1).

It has been our intention to achieve pulse powers in excess of 1 MW and Figs. 1 and 2 were obtained using eqn. (1) to see which parameters in our system should be changed to reach this power level. It is clear that shortening the pulse length but only allowing the oscillator to oscillate on one line is not very effective. Even shortening the pulse to 0·02 μs (the limit set by a 50 MHz line width) only produces an increase of two at $G/G_0 = 0{\cdot}5$ (Fig. 1). At 0·05 μs, the fastest rise-time we have observed with any of our lasers, the improvement at $G/G_0 = 0{\cdot}25$ only produces an increase of 50% over our present system (Fig. 2). Reducing q has a much more dramatic

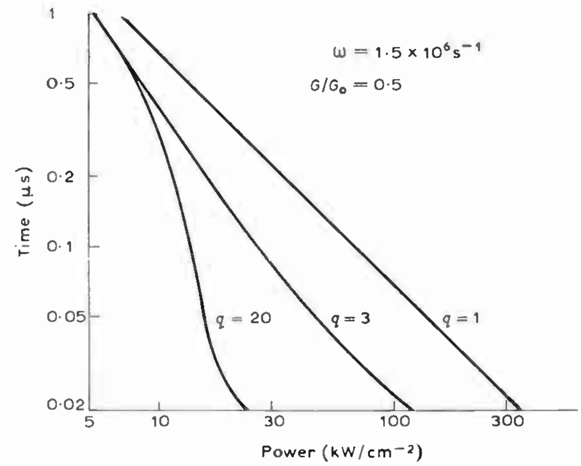


Fig. 1. Variation of power density obtainable as the pulse length is varied, for different q values (see text) at a constant saturation condition. (Theoretical results obtained using eqn. (1).)

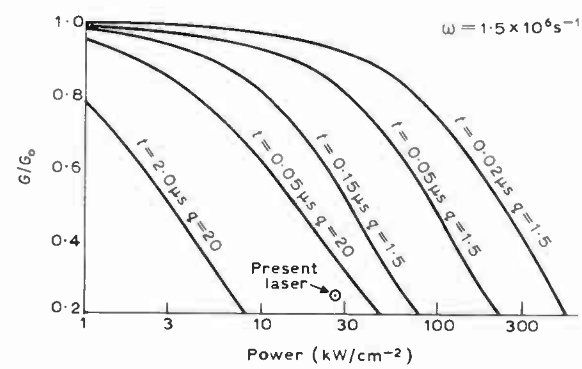


Fig. 2. Variation of power density obtainable as the degree of saturation (G/G_0) is increased, with different q values (see text) and different length pulses. (Theoretical results obtained using eqn. (1).)

effect, but not until q reaches about 3. (q is the ratio of the unused to used population of the rotational states in the oscillator.) We have now constructed oscillators which operate in axial modes but on many lines simultaneously and we estimate that the q value of these oscillators is 1·5–2. With low q values it is clear (Figs. 1 and 2) that not only is the saturation level increased by a factor of about 2·5 over the present system with the present pulse length of 0·15 μs but that shortening the pulse makes a very considerable further improvement. For example (Fig. 2), an amplifier with $\omega = 1{\cdot}5 \times 10^6\ s^{-1}$, $q = 1{\cdot}5$ and $t = 0{\cdot}15$ μs saturating to $G/G_0 = 0{\cdot}25$ would reach 65 kW cm^{-2} but shortening the pulse to $t = 0{\cdot}05$ μs would mean that we could reach nearly 200 kW cm^{-2}. With $t = 0{\cdot}02$ μs and $q = 1{\cdot}5$ the figure is 450 kW cm^{-2} and with $t = 0{\cdot}02$ μs and $q = 1{\cdot}0$ we reach 700 kW cm^{-2} or 10 MW in a 14 cm^2 area tube similar to our present system.

Realistic figures are t about 0·1 μs with $q = 1 \cdot 5$ and saturating powers of $\sim 100\ \mathrm{kW\ cm^2}$ which suggests that 1 MW pulses should be obtainable. The multi-line oscillators that we are using in the new system are narrow and have fairly low output powers. The a.c. drive method of the present system has proved rather restrictive so the new system will use narrow bore amplifier tubes d.c. driven which, because of the higher efficiency of narrow tubes, will actually give more gain per unit length than the original a.c. amplifiers. Assuming the final cross section of these amplifiers is 2·5–3 cm^2, power levels not much less than that obtained in the present system should be reached. The output at this level will then be passed through a 'Cassegrainian-type' optical system into a 5 cm diameter section also d.c. driven when well over 1 MW should be obtained.

Undoubtedly oscillation of the amplifiers will still be a problem but we have devised an optical component which assists in overcoming this effect. This device is termed a 'hole-blower' and consists of two off-axis paraboloids and two plane mirrors which bring the beam to an exact focus and then allow it to follow its original path. This can be placed between any two sections of the amplifier and a tape, opaque at 10·6 μm, is passed through the focus, thus separating the amplifiers. The rising edge of the amplified oscillator pulse 'blows a hole' through the tape and, as this only requires a small part of the pulse energy, the rest of the pulse passes through with little attenuation. Using 'triple play' recording tape this device works quite satisfactorily but only over a relatively restricted range of power levels. If the power is too small no hole is blown and if it is too large a plasma is formed which heavily attenuates the beam. Moving the tape out of focus allows a large hole to be blown without plasma formation but the absorption is still high, presumably because the 'debris' from the hole is still in the radiation path. Moreover, in a d.c.-excited system the tape must move fast enough for the hole to be clear of the focus before the amplifier's inverted population is sufficient to cause oscillation. The recovery time after a Q-switched pulse[1] is about 10^{-3} s in which time the tape has travelled 0·1 mm. This is sufficient movement in the focus, where the hole diameter is about 0·05 mm, but not out of focus, when the tape speed would have to be much higher. Under the best conditions pulses with rising edges of 0·25 μs have passed through the tape with an attenuation of less than 30%. While this is obviously not an ideal optical component it will produce a considerable improvement compared with our original system. Furthermore we now know[3] how to increase the saturating gain of amplifiers while reducing the small-signal gain and this will make a very significant improvement over the present system.

It has been assumed, when pointing out that short pulses of more than 1 MW are within the capability of CO_2 laser systems, that multi-wavelength operation is acceptable. This may often not be so and power levels likely to be reached on a single line are of interest. Reference to Fig. 2 indicates that the limit is now around 30–40 kW/cm^2 or about 0·5 MW in a 5 cm diameter tube. To improve the situation higher values of ω are required and we have, as yet, found no way of increasing ω much beyond $2 \times 10^6\ \mathrm{s^{-1}}$.

One of the uses for this high-power system is the production of plasmas.[4] This not only requires high-power pulses but also rather long ones, ideally 1–2 μs. Figure 2 shows how difficult it is to reach very high powers for long pulses and reducing q does not help because when $t > 1/\omega$ the rotational levels can be considered as a single entity. It is in practice rather difficult to produce uniform pulses of more than 1 μs in a Q-switched system. However, Gibson and Patel[5] have shown that it is possible to design a cavity which oscillates on a single line (by including a diffraction grating in the cavity) and 'feeds' in energy from the other rotational levels at just the right rate to produce long, approximately flat-topped pulses. The length of the pulse can be varied by changing ω. If these pulses can be amplified to a sufficient level (10–20 kW) they should be suitable for plasma production.

References

1. Crafer, R. C., Gibson, A. F., Kent, M. J. and Kimmitt, M. F., 'Time-dependent processes in CO_2 laser amplifiers', *Brit. J. Appl. Phys.* (*J. Phys. D.*), Ser. 2, **2**, pp. 183–95, February 1969. and
   Crafer, R. C., Gibson, A. F. and Kimmitt, M. F., 'Rate processes in the CO_2 laser'. Proceedings of Conference on Lasers and Opto-Electronics, I.E.R.E. Conference Proceedings No. 14, 1969.
2. Hotz, D. F. and Austin, J. W., 'Gain saturation flux and stimulated emission cross-section for the 10·6 μ line of CO_2', *Appl. Phys. Letters*, **11**, pp. 60–2, 1967.
3. Crafer, R. C., Gibson, A. F. and Kimmitt, M. F., 'Gain saturation in CO_2 laser amplifiers', *The Radio and Electronic Engineer*, **38**, No. 6, pp. 357–60, December 1969.
4. Gibson, A. F., Hughes, T. P., Kimmitt, M. F. and Hallin, R., 'Plasmas produced by focused CO_2 laser radiation', *Phys. Letters*, **27A**, pp. 470–71, 1968.
5. Gibson, A. F. and Patel, B. S., Private Communication (not published).

Manuscript received by the Institution on 25th February 1969. (Contribution No. 127/CC 60).

U.D.C. 535.3: 374.621.375.8

Gain Saturation in CO_2 Laser Amplifiers

By

R. C. CRAFER, B.Sc. †

Professor A. F. GIBSON,
B.Sc., Ph.D., A.Inst.P.‡

AND

M. F. KIMMITT,
B.Sc., A.Inst.P.‡

Reprinted from the Proceedings of the I.E.R.E. Conference on 'Lasers and Opto-Electronics' held at the University of Southampton on 25th to 28th March 1969.

Summary: Gain saturation in the CO_2 laser system presents a limitation on the amplification of Q-switched pulses and this has been studied in some detail. With pulses of around 0·5 μs saturation effects are observed at power densities of order $10^4\,W\,cm^{-2}$. The results of a theory of the time dependence of saturation are presented, together with supporting experimental data. This theory indicates that, at high power levels, the decrease in gain with time during the oscillator pulse can be described by the sum of two exponential decays. A value for the relaxation rate of the rotational sub-levels of the active vibrational states is obtained for various gas conditions. The implications of this relaxation rate on the design of high power oscillator/amplifier systems is discussed.

1. Introduction

Gain saturation sets a limit to the capability of a laser system. We have been concerned with the amplification of pulses derived from a Q-switched CO_2 laser oscillator. The amplifier gain at the beginning of the pulse equals the small-signal gain of the amplifier, which has been excited for the relatively long period between successive pulses. As, however, energy is extracted from the inverted population in the amplifier, the gain falls. This paper is concerned with the rate of decrease of gain due to saturation and the factors that determine it.

If, during the oscillator pulse, the gain decreases in a time comparable with the vibrational relaxation times or the $N_2 \rightarrow CO_2$ transfer time,[1] these relaxation processes influence the gain decay rate. At high powers, however, significant saturation occurs in times of order 100 ns and vibrational relaxation processes can be neglected. The only known processes likely to affect saturation rates in the sub-microsecond range are (a) relaxation between rotational levels and (b) hole burning within a given rotational line width. The latter is not important in CO_2 lasers at pressures above 10 torr,[2] and in any case not important for pulses derived from an oscillator using a rotating mirror Q-switch as the frequency of the emitted radiation is swept through the rotational line width several times during a single Q-switched pulse. There remains, therefore, relaxation between rotational levels. In the CO_2 laser inversion occurs between some 40 rotational sub-levels of the 001 and 100 vibrational states If the oscillator is constrained to operate on only one or a few of the possible transitions (P18, P20, P22, etc.), energy is initially extracted only from these rotational levels but may be replaced by relaxation from other, unused, rotational levels.

We have developed a theory of saturation on this basis, describing rotational relaxation by a single frequency ω. This work is described in more detail elsewhere.[3] If the energy stored in rotational levels *not* coupled to the radiation from the oscillator is large (i.e. the oscillator is operating on only 2 or 3 lines) we obtain for the gain as a function of time:

$$\frac{G}{G_0} = \left[\frac{P}{P+\omega}\right] \exp\left(-qt(P+\omega)\right) + \left[\frac{\omega}{P+\omega}\right] \exp\left(-\frac{P\omega t}{P+\omega}\right) \qquad \ldots\ldots(1)$$

where G_0 is the small signal gain, t is the time, P is proportional to the input power and q is the ratio of the total stored energy to the energy stored in the rotational levels coupled to the radiation. If q is less than about 3, eqn. (1) is invalid and no simple analytic solution is possible except in the special case $q = 1$ (all rotational levels used). When this condition applies, the decay of the gain is simply

$$\frac{G}{G_0} = \exp(-Pt) \qquad \ldots\ldots(2)$$

and independent of ω or q, as would be expected physically. The first term of eqn. (1), which is extremely rapid, arises from the direct extraction of energy from the selected rotational levels while the second, slower, term is primarily due to the feeding of energy into the selected levels from the uncoupled rotational levels.

2. Experimental Arrangement

Though measurements have been made on 1 in and 2 in (2·5 and 5 cm) diameter amplifiers, most quantitative data has been obtained using a 5 ft (1·52 m) long, $1\frac{1}{2}$ in (3·8 cm) diameter water-cooled

† Formerly at the University of Essex; now at the University of Hull.

‡ Department of Physics, University of Essex, Wivenhoe Park, Colchester, Essex.

tube with a gas flow rate of about 30 gas changes per minute. The choice of length was a compromise between sufficient gain for measurement and uniform power density along the tube. The signal was derived from an 18 ft (5·5 m) long Q-switched oscillator operating on the P20 line unless otherwise stated. The beam through the amplifier was defined by entrance and exit apertures, the latter being within the first Fresnel zone of the former and 2 mm diameter to select a nearly uniform intensity region. This arrangement was used to measure (a) the gain at a fixed time during the oscillator pulse and (b) the variation of gain with time during the pulse. An attempt was also made to observe the recovery of the gain after cessation of the pulse, as this represents a direct observation of rotational relaxation. To obtain this data a second, low power, Q-switched oscillator was set up so that the beams of both oscillators were coincident in the amplifier. The two oscillators were orthogonally polarized and an analyser at the amplifier exit allowed the low power pulses to be observed in the presence of the pulse causing saturation. In this way the gain after the primary pulse was sampled and gain recovery observed.

3. Experimental Results and Comparison with Theory

The quantity P in eqns. (1) and (2) is proportional to the incident power. The relationship is:[3]

$$P = \frac{2\omega_2}{I_s} \cdot I \qquad \ldots\ldots(3)$$

where ω_2 is the spontaneous relaxation rate between the 001 and 100 levels and I_s is the saturation power density under c.w. and homogeneously broadened conditions. The value of ω_2 is[1] $1{\cdot}1 \times 10^3\ \mathrm{s}^{-1}$ and we take $I_s = 22$ watt cm^{-2} from Hotz and Austin.[2] Hence P is known in terms of the power density. The only unknowns in eqn. (1) are ω and q, which may be deduced from the experimental data. We can, however, estimate q from its definition to be of order 20 and ω has been estimated theoretically by Tychinskii[4] to be of order $10^6\ \mathrm{s}^{-1}$. If these estimates are correct, the initial decay, represented by the first term in eqn. (1), is very rapid and the subsequent decay comparatively slow; representative time-constants are 50 ns and 5 µs, compared with a Q-switched pulse length$\sim$1 µs. Hence the gain at, say, 0·6 µs from the start of the pulse is practically independent of q, which appears only in the initial decay, and gain measurements at this time may be used to find ω alone.

We have measured the amplifier gain at 0·6 µs after the start of the laser pulse under various gas pressure conditions and with both d.c. and 50 Hz a.c. excitation. In Figs 1(a) and (b) we show the normalized gain (G/G_0) as a function of power for various helium

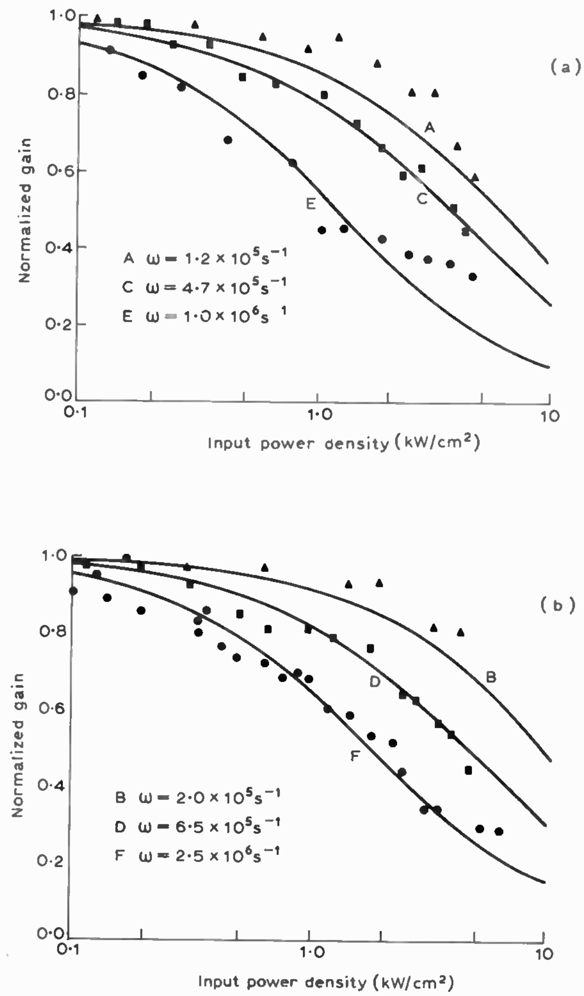


Fig. 1. Normalized gain as a function of input power density. Curves theoretical, points experimental.

pressures. The curves drawn are theoretical, based on numerical solution of eqn. (1) and eqn. (3) for the values of ω indicated. q was taken to be 21 but, as shown above, the curves are very insensitive to this choice. The agreement between theory and experiment is good for gas mixtures near optimum for small-signal gain (He pressures 4·2, 7·5 and 10 torr) and satisfactory for the remainder. Particularly noteworthy is the very slow decrease in gain with power density when compared with c.w. homogeneous saturation.[2] This is, of course, a consequence of energy being fed from unused rotational levels. It should also be noted that the theoretical curve for $\omega = 2 \times 10^6\ \mathrm{s}^{-1}$ differs only by about 20% from ω = infinity. Experimental points (neglecting scatter) beyond this curve would imply serious error in the absolute power measurements of eqn. (3). Hence the

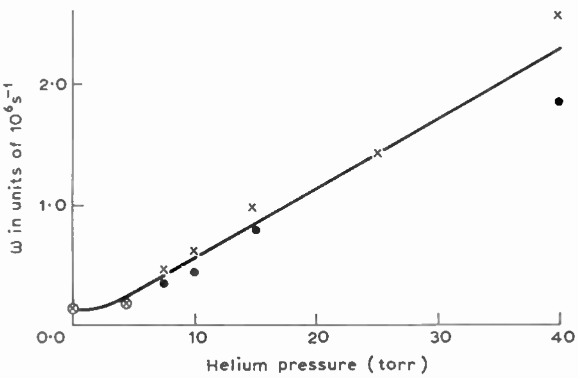


Fig. 2. Rotational relaxation, ω, as a function of helium pressure.

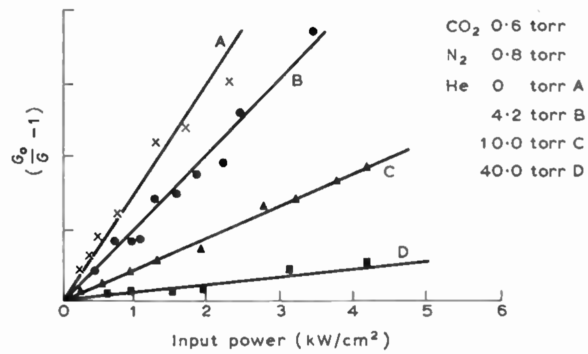


Fig. 3. Variations of $(G_0/G-1)$ with power from eqn. (6).

consistency of the data confirms, indirectly, the values of ω_2 and I_s, used in eqn. (3). This is further confirmed by Fig. 2, which shows that the deduced variation of ω with helium pressure is of the form expected theoretically: linearly increasing with pressure at high pressures but constant at low pressures.

Preliminary measurements of the decay of the gain during the oscillator pulse showed immediately a rapid initial fall (limited by the bandwidth of the detection and display equipment) followed by an almost constant gain plateau. Using the data of Fig. 2 we calculate that for an amplifier containing 0·6 torr CO_2, 0·8 torr N_2 and 4·2 torr He operated at a power level sufficient to reduce the gain by a factor 2 at 0·6 μs would show a slow decay-time constant of about 10 μs. So slow a decay would not be accurately observable during a 1 μs duration pulse. While making the observation of decay difficult (see below), this result suggested an alternative way of analysing the data of Fig. 1. If

$$qt(P+\omega) \gg 1 \qquad \text{......(4)}$$

and

$$\frac{P\omega t}{P+\omega} \ll 1 \qquad \text{......(5)}$$

eqn. (1) is reduced to the very simple form

$$\frac{G_0}{G} - 1 = \frac{P}{\omega} \qquad \text{......(6)}$$

so that $(G_0/G-1)$ plotted against P should yield straight lines through the origin. Such plots are shown in Fig. 3. Values of ω derived from the slopes of these lines are included with those from Fig. 1 in Fig. 2. That the former values are consistently equal or lower appears to be due to the inadequacy of approximation (5) at large ω.

It is clear from the above that to observe even the slow component of the gain decay directly, the following conditions must be achieved simultaneously:

(1) a large value of ω by addition of helium to the amplifier;

(2) very high power and approximately 'flat-topped' pulses from the oscillator.

These conditions proved difficult to achieve and measurement accuracy was low as high powers and high ω imply very small gains. Some representative decay curves are shown in Fig. 4. The results refer to a 0·6, 0·8, 18 torr mixture for which ω (from Fig. 3) is $1{\cdot}2\times10^6\ s^{-1}$. Every curve drawn in Fig. 4 is theoretical, based on this value of ω and the power level stated. The agreement between theory and experiment is satisfactory.

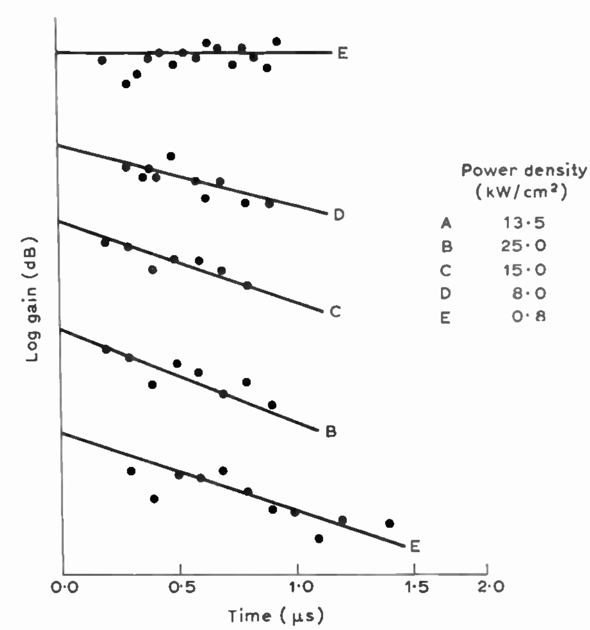


Fig. 4. Logarithm of gain in dB versus time during laser oscillator pulse for $\omega = 1{\cdot}2\times10^{-6}\ s^{-1}$. (Curves displaced for clarity).

The rate of recovery of the gain due to rotational relaxation may be deduced from eqn. (1) when $P = 0$ and is:

$$G = G_0[1-\exp(-q\omega t)].$$

So the recovery is strongly dependent on q while the gain throughout most of the pulse is substantially independent of q. Thus the theory predicts that the amplifier performance will be a function of the rotational line spectrum of the high power oscillator when this has switched off, but not during most of the time it is on. This prediction provides a direct check that the gain saturation is limited by rotational relaxation. The recovery is, however, extremely rapid and difficult to measure unless q is reduced by increasing the number of lines upon which the oscillator is operating.

Recovery rates have been measured with the oscillator operating on 3, 4, 5 and 6 lines (P16 to P26 inclusive). The recovery rate is observed to decrease with increasing lines, as expected, and the value of q estimated from these data is 30 ± 10. Thus the recovery rates are of the expected order of magnitude.

4. Conclusions

Notwithstanding the simplification implied by describing the relaxation of the rotational levels of the 001 and 100 vibrational states of CO_2 by a single rate, ω, the model adopted gives an adequate description of our experimental results. The value of ω increases with N_2 and He pressure as anticipated by Tychinskii[4] though the absolute values are somewhat lower than his estimate. We have also found that water vapour at comparatively low pressures can increase the value of ω significantly.

The optimum conditions for an amplifier operating near saturation level depends on the objective to be achieved. Maximum efficiency of extraction of stored energy by a pulse from a Q-switched oscillator occurs when the oscillator is extracting energy from all rotational levels simultaneously (i.e. $q = 1$). Under these conditions the rotational relaxation rate is of no significance. If, on the other hand, spectral purity is required and the oscillator must be limited to one rotational transition, ω should be increased to permit utilization of the energy stored in otherwise unused levels. Increasing ω by the addition of N_2 or He reduces the small signal gain so there is an optimum compromise, depending on the desired result. A possible figure of merit is

$$G_0 P_{\frac{1}{2}}(t)$$

where G_0 is the small signal gain and $P_{\frac{1}{2}}(t)$ the power at which the gain is reduced by a factor of 2 at a time t after the start of the pulse. For $t = 0{\cdot}6$ μs helium at a pressure between 15 torr and 20 torr gives the highest figure of merit we have observed. The performance at other times and pressures can, of course, be computed from eqn. (1).

5. References

1. Crafer, R. C., Gibson, A. F., Kent, M. J. and Kimmitt, M. F., 'Time-dependent processes in CO_2 laser amplifiers', *Brit. J. Appl. Phys. (J. Phys. D.)*, Ser. 2, **2**, pp. 183–95, February 1969, and
   Crafer, R. C., Gibson, A. F., and Kimmitt, M. F., 'Rate processes in the CO_2 laser', Proceedings of the Conference on Lasers and Opto-Electronics, I.E.R.E. Conference Proceedings No. 14, 1969.
2. Hotz, D. F. and Austin, J. W., 'Gain saturation flux and stimulated emission cross-section for the 10·6 μ line of CO_2', *Appl. Phys. Letters*, **11**, pp. 60–2, 1967.
3. Crafer, R. C., Gibson, A. F. and Kimmitt, M. F., 'Rotational relaxation and gain saturation in CO_2 laser amplifiers', *Brit. J. Appl. Phys. (J. Phys. D.)*, Ser. 2, **2**, pp. 1135–46, August 1969.
4. Tychinskii, V. P., 'CO_2 lasers', *Soviet Physics Uspekhi*, **10**, pp. 134–52, 1967.

Manuscript received by the Institution on 25th February 1969 (Contribution No. 128/CC61).

U.D.C. 656.25:681.3

A Computer Controlled Train Describer

By

L. J. BENTAL,
B.Sc., M.Eng.†

G. S. EVANS,
C.Eng., F.I.E.R.E.‡

AND

A. J. TOMKINS,
M.A., C.Eng., M.I.E.R.E.‡

Presented at a meeting of the Joint I.E.R.E. and I.E.E. Computer Groups held in London on 25th September, 1968.

Summary: In modern railway signalling practice each train is allocated a code which describes its type, priority and route. The code is displayed on a schematic display diagram in the signal box and enables the signalman to set up the correct route for the train to follow. The description code on the diagram automatically follows the movements of the train, and in addition, is transmitted to the neighbouring signal box when the train enters the adjoining area. The equipment used to display and manipulate these codes is known as a train describer, and the paper describes the first computer-based system to be brought into operation.

1. Introduction

The absolute Block System of train control and signalling is now mandatory for mainline railways in the United Kingdom and ensures separation of consecutive trains by dividing the track into sections, on which normally only one train is allowed at a time. Such block sections were originally controlled by individual signal boxes in each of which was vested complete responsibility for accepting or rejecting a train offered from a previous section.

The development of the electrical 'track circuit' allowed the detection of the presence of a train on a particular track section. This permitted the introduction of rudimentary mimic diagram displays in the signal box, on which lamps were geographically placed and then illuminated automatically to show the progress of a train over a very limited length of track. Figure 1 shows the operating floor of a traditional signal box of this type.

With the advent of colour light signals instead of semaphores and power-operated points instead of mechanical linkages, it has become possible to control wide areas of track from a remote central point. This centralization ensures more efficient control as a wider view of the traffic situation may be readily observed by the signalman. There is also a striking reduction in staff, and in the necessary communication network between signal boxes as the very numerous traditional signal boxes are increasingly replaced by a much smaller number of large power operated boxes.

In a large power signal box with control of many trains at one time, a code showing the identity, priority and routing information for each must be displayed to the signalman. This code, known as the headcode or train description, normally consists of four characters, the second character being alphabetic whilst the remainder are numerals.

Figure 2 shows the operating floor of a modern signal box. Signals and points are operated by push buttons which replace the heavy mechanical levers, and a sophisticated mimic diagram provides full information about the states of all blocks within the area in three ways:

(i) Routes which have been set up are indicated by strings of white lights.

(ii) The presence of trains is indicated by moving strings of red lights.

(iii) The identification of a particular train is indicated by displaying its code on geographically-placed small cathode-ray tubes.

The display of the train code on the diagram and its automatic movement in accordance with the movement of the train on the track requires its own control system, commonly known as a train describer.

Because operation times are relatively slow, control and signalling equipment for railways traditionally has been constructed of relays. This is still true of the safety interlocking circuits where the specially produced fail-safe relay cannot yet be superseded. However, in equipment not directly concerned with the safety of trains the relay is now being replaced by solid-state equipment with consequent savings in space and maintenance requirements.

Previously, electro-mechanical and solid-state train describers were specially designed for each application.[1] This necessarily entailed a great deal of special engineering work for each installation and, although standardization was pursued, the whole equipment was specifically designed to meet the very limited market for train describers alone.

For an equipment to be produced and installed at Leeds for British Railways it was resolved to use a

† Marconi-Elliott Computer Systems Limited, Borehamwood, Herts.

‡ Formerly with Elliott Automation Systems Limited; now with Warren Point Limited, Welwyn, Herts.

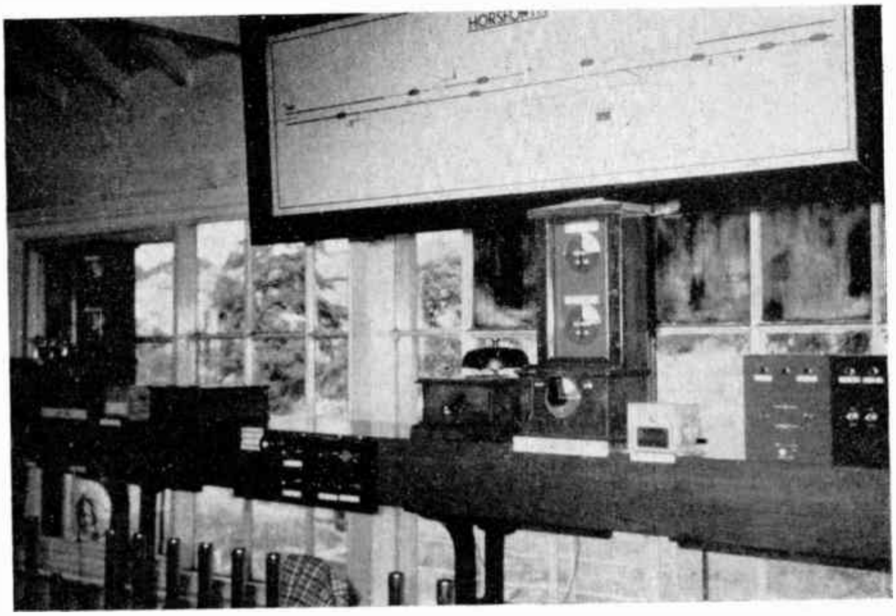


Fig. 1. Operating floor of traditional signal box.

standard general-purpose computer, together with as much suitable standard peripheral hardware as was already in use across many fields of application. Such additional and specifically railway-orientated units as were required were designed to be suitable for repetitive production and generalized application.

General-purpose computers are now mass produced. This has resulted in a considerable reduction in price so that it is now economical to use general-purpose computers in installations where previously it was not practical. In addition, testing procedures have been standardized and wide and efficient maintenance facilities are available.

Furthermore, in a general-purpose computer the system requirements of a particular installation are implemented by program software. This means that the installation can be adapted to future requirements by changing the software only. The new program can be tested out beforehand and fed into the computer on site within a matter of minutes. In addition, the use of software control reduces the amount of equipment required, not only reducing the physical size of the installation but also increasing its reliability.

2. System Requirements

The permanent way and signalling re-organization in the Leeds area of British Railways has resulted in the central control of over 50 route miles of track, with a radius of influence of about 20 miles, centred on Leeds. A new five-storey building in the station precincts contains the equipment and administrative facilities for the area. Seven conventional mechanical signal boxes have been selected to define the limits of the control area, and these are required to communicate over land lines with Leeds, thereby behaving as interfaces between the modern central control system and the conventional signalling network.

Planned expansion of the scheme will result in the closure of the present outpost boxes and the transfer of their responsibilities to Leeds central control. More remote boxes will then assume the interface role. Thus the complexity of the mimic diagram at Leeds will increase, and the number of cathode-ray tube displays rise from an initial 90 to 800.

The basic system requirements for the train describer may be summarized as follows. The descriptions of trains leaving Leeds station are inserted by the signalman by means of a set-up panel on his operating console. It is necessary, therefore, for the apparatus to assemble a four character headcode description of a train, in response to the signalman pressing the relevant buttons on a keyboard. The system then inserts the assembled code into the appropriate signal berth on the diagram selected by the signalman from his knowledge of the route the train will take. The stored description is then continuously regenerated in order to present an apparently static display on the associated cathode-ray tube.

As the train moves along the track, electrical impulses derived from the signal and track circuit relays cause the description to step along the diagram in sequence with the train. All the trains originating in Leeds are inserted on the diagram in this way.

Those trains starting outside the area and proceeding to Leeds are set-up in the manual-type signal boxes on the fringe of the area and are transmitted to the train describer equipment, where they are received and displayed automatically in the correct berth on the diagram.

3. System Design

The essential elements of the system that was finally developed are shown in Fig. 3. It may be considered in four main parts.

3.1. ARCH

A feature of the majority of on-line computer systems is a requirement to read sequentially a large number of inputs, and to output information to a large number of destinations. Since all computers have limited direct input/output facilities it is necessary to construct a multiplexer system to allow the required inputs or outputs to be gated on to the highways.[1] It was decided to use a standard hierarchical system of multiplexing equipment produced by our Company under the generic name of ARCH which has been available for several years, and is well described elsewhere.[2]

Briefly this is a system of modular units which can be connected together in the required combination to cater for the inputs and outputs required for any system, in this case approximately 300. Various types of analogue and digital input and output interfaces are available, although in the train describer application only digital units are required. The ARCH equipment also contains a system alarm unit which, unless reset by the computer program at frequent intervals, will activate a visual and audible alarm, thus maintaining a continuous check on correct operation of the system.

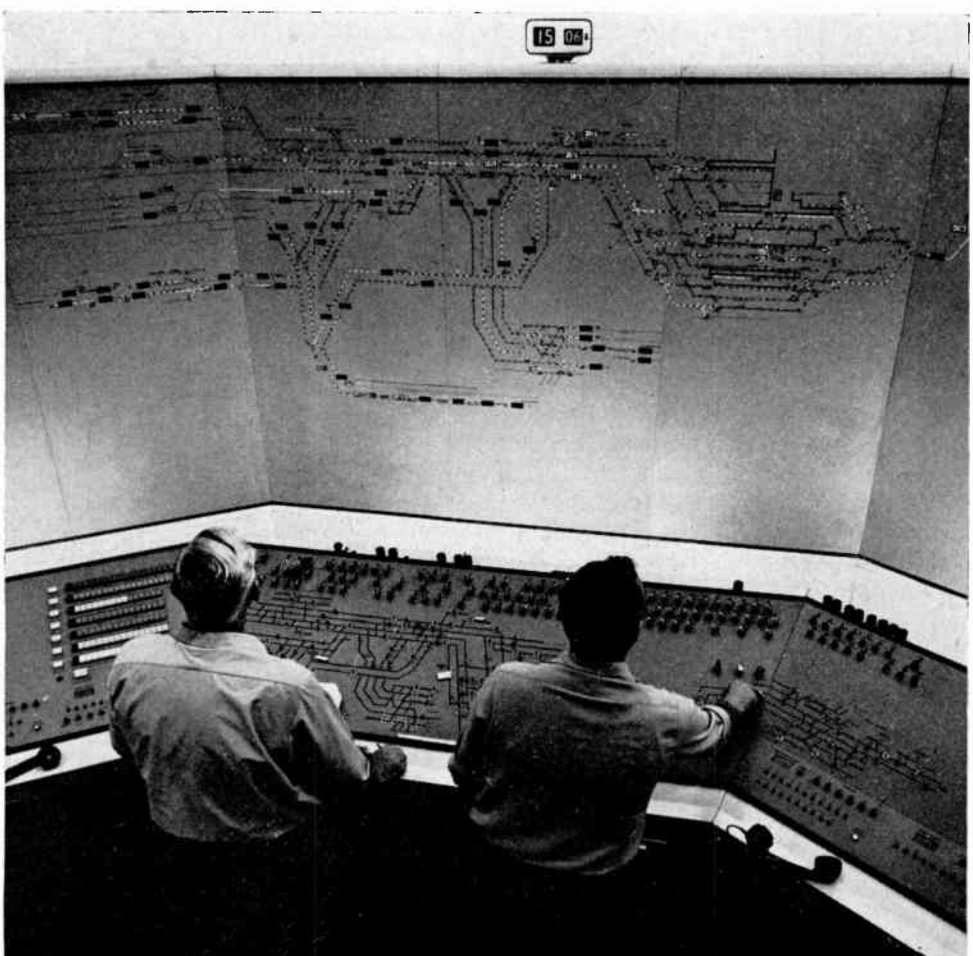


Fig. 2. Leeds signal box operating floor.

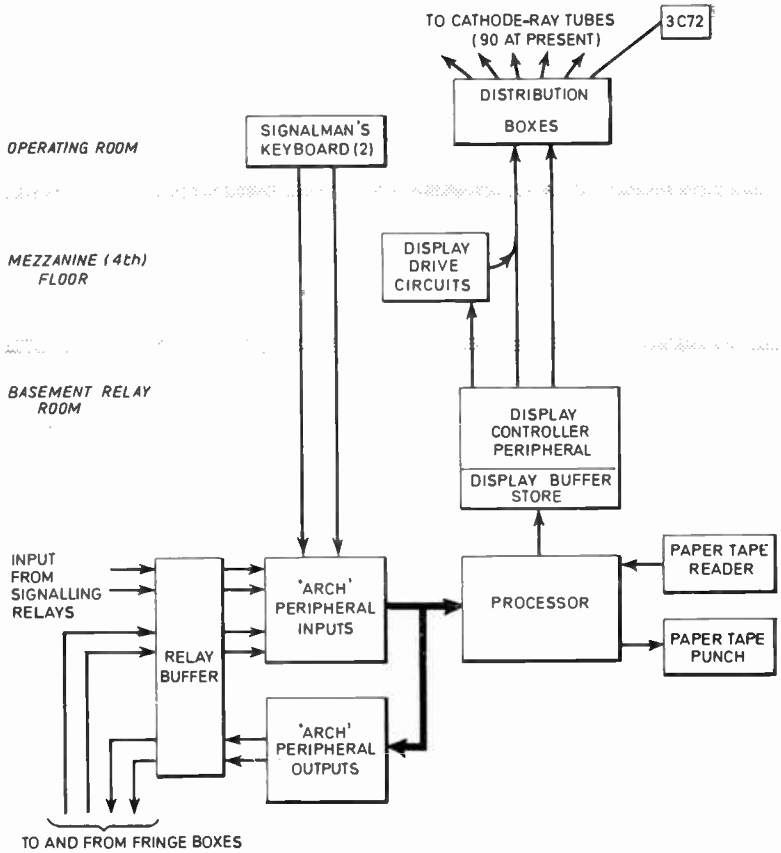


Fig. 3. Leeds train describer block schematic.

3.2. *The Computer and Display Store*

The train describer system is based upon the standard Elliott 903, which is a fully parallel computer with a word length of 18 bits and an internal 8192 word store of 6 μs cycle time.[3] Operation times are relatively long in comparison with later computers of the 900 series, but are entirely adequate for this application, as the present workload of the computer has been found to occupy less than 25% of the available time.

A further 8192-word ferrite core store unit is included in the system to contain the information actually displayed at any instant. This store is continuously accessed by the display controller circuits in order to generate the required displays and is also accessed by the computer when a change in the displayed information is required, following an external demand on the system. This store is identical to that in the computer and allows the provision of up to 200 displays.

Both the 903 computer and its auxiliary store are equipments which have been produced in very large quantities, and as such are subjected to stringent quality control and manufacturing inspection. They are commercially packaged versions of the 920 computer which was designed primarily for ruggedized use in a military environment, and are thus very suitable for applications such as that here described where reliability is of prime importance.

3.3. *The Display Controller*

The display system is arranged so that the cathode-ray tubes receive parallel deflection signals, and suitable staircase wave-forms are supplied simultaneously to the X and Y deflection plates of all cathode ray tubes. Alpha-numeric characters are displayed as matrices of 7 × 5 dots by providing bright-up pulses on the cathode of a particular tube in the correct time sequence, as the electron beams scan the complete raster.

The data for both the generation of the deflection wave-form and the bright-up information for each tube are contained within the display store, which ensures that the display bright-up information cannot become out of phase with the deflection signals. The display controller is designed to extract this information and generate the required signals to control the c.r.t.s.

The controller is constructed using standard diode-transistor logic, identical with that used for the computer and auxiliary store. It is designed to be very readily expandable, and all control logic is designed to operate satisfactorily a total of 800 displays, which is thought to be the probable upper limit of the requirement for power signal boxes.

3.4. *Display Drive System*

It was appreciated from the start that the quality of the visual display was a prime consideration for the train describer, since this is the end-product of the whole system.

At Leeds, the 78 m (85 yards) separation between the display controller and the cathode-ray tubes clearly presented a severe problem in driving the latter so as to provide a clear and undistorted display. The standard of display which was finally achieved is shown in Fig. 4.

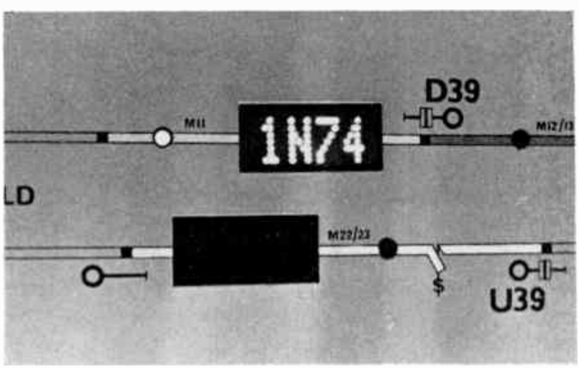


Fig. 4. A sample display.

The X and Y deflection amplifiers are fitted in a cubicle adjacent to the display diagram, and special circuits have been provided to allow the transmission of the deflection signals from the display controller to these amplifiers. To prevent possible distortions, the signals are transmitted in binary form over a coaxial cable which has circuits giving high noise immunity at either end. The staircase wave-forms are then produced in digital-to-analogue converters which form the first stages of the amplifiers. After phase-splitting to produce the two antiphase deflection signals required, the entirely solid-state amplifiers develop some 200 W total output power. This power is necessary to overcome capacitance effects in the long runs of screened cables to the c.r.t.s which are still about 9 m distant.

The bright-up signals for the tubes are also transmitted in binary form from the display controller via small line driving amplifiers and coaxial cables. They are routed, as are the amplified deflection signals, through distribution boxes which can each supply up to eight tubes. Individual brightness adjustment is provided for each tube and a control on the signalman's control desk allows variation of the overall brightness of the complete display system by changing the length of the bright-up pulses applied to the tube grids.

Circuits have been fitted which ensure that a total failure of the deflection amplifier system will cause the e.h.t. supply to be removed, so that screen burn is not caused by the undeflected spot. The design is also such that in the event of most deflection amplifier failures, the display will remain legible, although distorted, until the fault can be repaired.

3.5. *Main Box to Outpost Box Communication*

Data must be transferred between main box and outpost boxes as trains pass from the control area of one signal box to another. The transmission is automatic when a train leaves the main box area, but where a train is travelling towards Leeds, the signalman at the outpost box initiates a transmission by push button. However, the data transfer is identical in either case.

The transmission is a 21-bit binary coded message, with a parity bit included in each of the four character codes. A message is transmitted in a three-state self-clocking fail-safe mode, and to preserve electrical isolation, relays are used as line drivers.

At a receiver, the message is stored and automatic checks of its integrity carried out. If all conditions are fulfilled a check-back pulse is returned to the transmitter, otherwise a second transmission is initiated automatically. Should another failure occur, alarms are given at both main and outpost boxes and the signalman may resort to telephone communication until the fault in the transmitter system is cleared.

In the main box, checks and double transmissions are made directly by the computer, which also determines the timing of the pulse trains. It is a significant advantage of the computer-based system that the form of the transmitted message and its timing are readily changed by program to provide a different interface with an adjacent train describer system. In the Leeds equipment there are a total of 10 two-way transmission links and it is possible for all these units to be in action at once.

4. System Program Design

When an on-line process is controlled by a general-purpose computer, the processor is required to produce certain outputs in response to the state of a particular input from the real-time system. The response that must be generated is determined by the program, and may depend upon stored tabular-type data and on such factors as the previous state of the inputs and the actions already performed. The program, which will normally control the sampling of the inputs as well as details of the necessary actions to be performed, is normally placed into the computer's internal store by means of punched paper tape or a similar medium.

It is almost invariably true that the computer will only deal with one action at any instant, but the speed of a digital computer is such that the sequential implementation of many hundreds of actions may be carried out with the appearance of simultaneity.

4.1. *Levels of Priority*

Operations in on-line control systems fall normally into definite priority groups. High priority contains the supervisory program, medium priority levels determine the actions to be taken, whilst the low priority levels actually execute the actions.

In a train describer system, the highest priority level contains the overall supervisory, initiating and manual control programs.

The medium priority level is entered every 33 milliseconds in response to an interrupt signal generated by the ARCH hardware. It causes the scanning of the track relay contacts, manual inputs and incoming data lines to detect any change of state which has occurred. In addition this level sets all outgoing data lines.

The low priority level comprises the programs which prepare and execute the actions determined by the medium priority level, and contains the main workload of the computer. This includes analysing changes of state in track and signal relays, setting up new descriptions, moving descriptions on the display, decoding and checking of incoming messages and preparation of outgoing information. This level also contains self-checking procedures which check automatically that the computer and associated equipment function correctly.

4.2. *Stacks*

An important feature of the program is the use of special lists, called stacks. While the computer has ample time to deal with the average work load of the train describer, it is still possible for demands for action to occur simultaneously. In order, therefore, to avoid the possibility of missing one action, while the computer is obeying another, actions are queued in a stack. When an action is demanded, a code number is placed in a stack, and these numbers are later extracted from it and obeyed in turn, on the basis of the oldest item first.

This is exactly analogous to an air traffic controller who can only deal with one aircraft at a time. Where the rate of approach of aircraft exceeds the resources of the system, he will acknowledge and note their arrival and relegate them to a stack until he is ready to deal with the earliest arrival.

4.3. *Checking Program*

The program carries out certain checks during the course of its normal work. In addition, there is a set of special self-checking procedures, designed to show up any fault in the computer itself, or corruption of its data. The self-checking routine is entered when no other work is required, and the stacks are empty. This is, at the present, the case of approximately 75% of the time. The self-checking routine is split up into short blocks and after completing each block, the computer tests whether any other actions are required. This ensures that there is no significant delay in carrying out actions, even though the complete checking cycle may take several seconds, particularly if discrepancies are found.

Apart from checking that the computer is working properly, the program also contains routines for:

(a) Sum-checking the program itself.

(b) Cross-checking of the two duplicate sets of current description lists contained in the computer.

(c) Checking the display auxiliary store contents against the description lists held in the computer.

If any error is found, error indications are given by lamps on a special panel, or by an output on the computer's tape punch, as appropriate. This self-checking facility is a very valuable aid to reliability. For example, if it is found that a description has been corrupted by one or two spots only, the computer will automatically restore the data so as to display the correct description. Whilst it is true that a train describer system using special and expensive hardware can be engineered to produce error indications in the event of a failure, it is contended that only by the use of a general-purpose computer-based system can continuous checks be carried out economically to ensure that the system is still functioning normally. Further, by the use of marginal testing techniques it is possible to detect areas liable to become faulty and to correct these before interruption in service occurs.

5. Engineering Problems

5.1. *Interfacing with the Railway Signalling System*

One of the problems associated with the introduction of electronic equipment into railway signalling is the difficulty of interfacing very high speed electronic equipment with slow and operationally loosely defined signalling relays, some of which may have been installed over 70 years ago. These relays are large devices specially designed to fail-safe in all aspects. They are therefore slow to operate, and when approaching the end of their working life may bounce for more than 100 ms after operation.

Also British Rail insist that any connection to a signalling relay must be earth-free, since stray earth currents might, in some extreme circumstances, result in the faulty operation of other signalling relays and the consequent possibility of a malfunction of the signalling system.

The initiation of a movement of a description on the display panel is derived from the signalling equipment as a closing relay contact. This closure is made to operate a slugged reed relay in the train describer, thereby overcoming both problems of bounce and isolation. The output of the reed relay is used as an input to the ARCH multiplexer and when its closure is detected by the computer, the necessary program action is initiated. Since the reed relay output is isolated from the signalling system relays, it is permissible to introduce at this point the earth reference necessary for the electronics.

By the use of these techniques and program controlled sampling, it has proved possible to maintain a completely reliable operation of the interface between the signalling system and the electronic train describer.

5.2. *Reliability Consideration*

One of the primary requirements of railway signalling equipment is that it should be reliable as regards both its integrity and its service availability. Although train describer equipment is not directly concerned with safety aspects of the signalling system, it has become vital to signalmen working complicated junctions, and its operation is essential for the unrestricted running of a closely-spaced timetable.

Because of this importance, and the fact that in order to achieve high utilization the railway is often used as heavily at night as during the day, the train describer is required to be fully operational for twenty-four hours every day. Engineering work is sometimes possible for short periods during Sundays, but even this time is very limited.

Whilst solid-state equipment is inherently reliable, this is only true if correct design philosophies are pursued. In the design of the train describer silicon semiconductor devices were used throughout, together with professional-grade capacitors and metal oxide resistors. Circuits were designed so that maximum stress levels of 10% for each component were not exceeded and all designs were toleranced for worst case conditions. By the use of these techniques, properly applied, a very reliable equipment was produced.

However, probably the most significant factor in achieving reliability was the extensive use of standard units which were in high quantity production. Use of this type of equipment, mainly designed to military specification, allows quality control of design and manufacture almost impossible to obtain under small quantity special-purpose production. It also reduces the inherent unknown factors in the engineering design and allows a more extensive maintenance back-up service.

It was not felt that formal reliability calculations would be meaningful due to the statistical nature of the methods used, and the fact that the train describer system would be a unique equipment. This was born out by the fact that, although calculations based upon component counts and standard failure rates led to an m.t.b.f. figure of only 1000 hours, the system has now been in continuous use for over fifteen months (equivalent to 10 000 hours) without a single central equipment failure. This performance gives good support to the design philosophy adopted for reliability.

6. Future Developments

Probably the first and most natural extension of the system will be to produce automatic records of train running. At present, the arrival, departure and presence of trains at reporting points is usually recorded manually, by an operator specially employed for the purpose. These records could be produced extremely conveniently and cheaply on the tape punch associated with the computer. Such a tape record could be processed off-line to sort the data into whatever configuration is required, an on-line typewriter or line printer could easily be fitted, or graphs of train running could be produced directly using an on-line graph plotter.

Automatic announcing of the arrival and departure of trains at stations is another possible extension of the system, with the messages being pre-recorded on a tape-recorder controlled by the train describer. However, a difficulty in this application is caused by the often unpredictable nature of the required messages. Nevertheless, the system could no doubt be useful under certain circumstances, such as in the case of small stations some miles from the signal box where messages are likely to be relatively simple and few in number.

To look further ahead, an obvious application for the train describer is for the automatic routing of trains through the system, based on the head-codes carried by the trains. Technically, this does not appear to be a particularly difficult problem, but the difficulties encountered are likely to be more of a cybernetic than practical nature.

For example, it is by no means certain that the information contained in the standard four-digit head-code used by British Rail would be sufficient to provide all the routing data required. However, it is a particularly valuable asset of the computer-controlled train describer that it would be very easy to increase the size or nature of the description held by the computer, even though the head-code displayed to the signalman could remain as at present. In this way it would be possible to let the describer use information additional to the standard head-code for the purpose of automatic train routing.

Studies of the effect of priorities of trains in a complicated junction are as yet by no means complete, and the conditions that the routing must fulfil are, therefore, still to be accurately defined.

This is by no means a simple matter due to the very widely differing speeds of trains of different priority, but it is felt that the routing of trains through a simple diverging junction could readily be achieved in the near future. However, British Rail are justifiably proud of their safety record, and the primary requirement of all railway signalling equipment is that it shall be proved to be extremely reliable. Electronics in signalling is very much in its infancy, and will be required to show its fulfilment of the above conditions before any great strides are made in its incorporation directly into the signalling system, but there can be no doubt that, given time, computer-controlled equipment will be shown to be adequate, and will be used for the direct automatic control of railways.

It has been suggested that the ultimate railway control system may consist of a small computer-controlled train routing system in each power box, almost certainly based on the train describer, which will then be in continuous automatic communication with a larger computer in a control centre, which is deciding the control strategy for that area.

This is not to imply that the computer system could be run without human aid, but it is almost certain that the routine tasks at present performed by signalmen could be taken over by automatic equipment, leaving the operators to intervene only when very complicated dislocation occurs.

In conclusion, perhaps it is right to temper our enthusiasm by the knowledge that such a philosophy is unlikely to be widely adopted within the next decade but it is probably the ideal system for the future.

7. Acknowledgments

The help and encouragement given in this project by Mr. J. E. H. Tyler, Chief S&T Engineer, British Railways Board, and Mr. R. E. Green, lately Chief S&T Engineer, Eastern Region, is gratefully acknowledged.

The authors wish to thank the Boards of G.E.C.-Elliott Traffic Automation Ltd. and Marconi-Elliott Computer Systems Ltd. for permission to present and publish this paper.

8. References

1. Peat, R., 'Electronic train describers', *Proc. Instn Railway Signal Engrs*, 1968–69. (To be published).
2. Cole, G. B. and Clarke, S. L. H., 'The development of ARCH—a hybrid analogue/digital system of computers for industrial control', *J. Brit. Instn Radio Engrs*, 26, No. 1, pp. 45–48, July 1963.
3. '903 Manual', Vols. 1–4, Computing Division (Borehamwood), Marconi-Elliott Computer Systems Limited.

Manuscript first received by the Institution on 28th January 1969 and in revised form on 18th August 1969. (Paper No. 1293/Comp. 124.)

Points from the Discussion on Program Design

Mr. M. Parker: It was known at the beginning of the project that the Leeds train describer system would be expanded from the original 80 up to 800 displays. With this in mind all programs were written so that any extension could be incorporated into the system with the minimum of program change. As all the description manipulation is done by the program, a complete new system program can be compiled and tested on another computer off-line without having to take the train describer out of service. The new tape can be read into the computer within seconds and the system tested on-line. If errors are detected then the old system program can be re-input while the new tape is being corrected so allowing the describer to remain operational.

The extension of the system to Garforth during 1968 proved that comprehensive additions to the system program could be made up on to paper tape, tested on another computer and finally tested on-line with the minimum of interference to the normal running of the train describer system. In fact, the new program was tested on-line in about two hours (during which time the original part of the system was functioning normally).

Another important feature of the program design is its ability to detect corruption within itself. Because the normal workload on the computer takes 25% of the computer time it was decided to spend the remainder of the time in self-checking programs.

The self-checking programs take about 6 seconds to complete during the normal running of the system, and if corruption of the display information held in the auxiliary store is detected, the pattern is built up again so that the distorted picture will be visible for only a few seconds.

The programs are written in SIR (Symbolic Input Routine), a simple machine code orientated language. The train describer program is held as a sum checked binary tape and is read into the computer under initial instructions.

U.D.C.535.4:681.327.57

Design of a Large Read-only Holographic Memory

By

R. M. LANGDON,
Ph.D.†

Reprinted from the Proceedings of the I.E.R.E. Conference on Lasers and Opto-Electronics held at the University of Southampton on 25th to 28th March 1969.

Summary: Factors affecting the design of a read-only optical memory having a total capacity of the order of 10^8 bits with fast random access are discussed. Photographic information storage is used and random address selection is provided by a laser beam scanned across the photographic plate using non-mechanical beam deflection. 'Page' organization of the memory involving parallel read-out of 10^4 bits from each of 10^4 beam addresses is used in order to achieve the large total storage capacity. The information array from each address is projected on to an array of 10^4 photodetectors and selection of the required bits from each 'page' is done electronically. Information is stored at each address in the form of a small hologram, which offers the following advantages over the conventional method of storage: uncritical alignment of the photographic plate, reduced sensitivity of the plate to accidental damage by incorporating redundancy in the hologram, and the ease with which the reconstructions of each hologram may be superimposed at the output photodetectors.

1. Introduction

The read-only memory, which can be defined as a memory in which the stored information is not alterable at electronic speeds, can be used for a number of purposes in digital computers. For example, it can be used for the storage of tables of mathematical functions, as a code converter for translating programming instructions into directions for opening and closing gates throughout the machine and for special purposes where translation from one digital code into another is required for example in a language translation machine.

To compete with existing techniques, such memories should have random access times of a microsecond or less and have a very large capacity ($> 10^6$ bits) at low cost per bit. There is a possibility that the optical memory using the photographic plate as a storage medium may be competitive with existing systems in applications requiring very high storage density.

This paper considers the design of a page-organized memory using holographic storage of the type described by Smits and Gallaher[1] and reports the results of some relevant experiments.

2. Design of the Page-organized Holographic Memory

The principles of page-organization are illustrated in Fig. 1.

This type of memory involves the storage of a number of bits at each beam address which can be read-out in parallel by focusing them on to an array of photodetectors. Further selection of the required bits is done by addressing the outputs of the photodetectors by normal electronic means. There are potential advantages in storing the pages in the form of holograms. These are considered below.

The optical system proposed for recording and read-out of page holograms is shown in Fig. 2. The 'page' of information to be recorded is first of all made into a transparency of black and white dots. This could be done, for example, by photographing the array of information displayed on the face of a cathode ray tube. This transparency is illuminated by a plane wave from a laser and the transmitted light is focused by the first lens L1, (the object lens), on to the hologram plate. By placing the transparency in the front focal plane of L1, transmitted light from every point on the object is converted into a plane wave as it passes through the lens. A plane reference wave is also incident on the plate and interference between this and light from each object point creates a transmission function in the hologram having a single spatial frequency covering the whole area of the hologram. When the developed hologram is illuminated by a read-out beam having the same direction of propagation as the reference beam, the first-order diffracted beam produced by the hologram is an array of plane waves having the same direction of propagation as the light from the original object points. When this is focused by the lens L2, an array of light spots is produced in the image plane corresponding to the points in the original object.

In the optical memory a number of such 'page' holograms are placed side-by-side and are addressed by directing the read-out beam on to the appropriate hologram. Provided that the angle of incidence of the

† Applied Physics Group, The Marconi Company Limited, Great Baddow, Chelmsford, Essex.

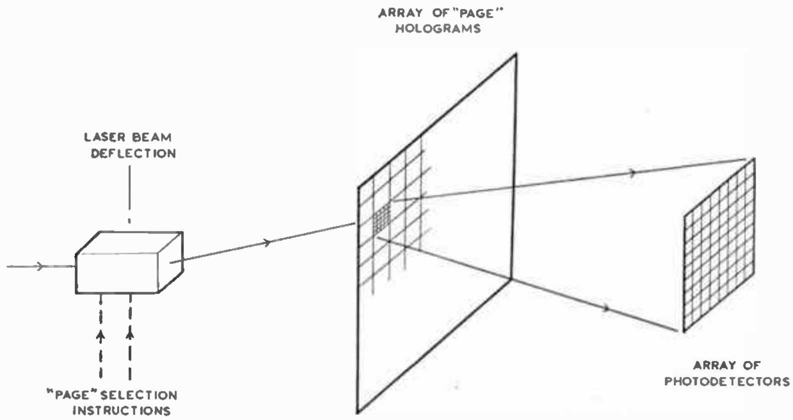


Fig. 1. 'Page' organized optical memory.

read-out beam is identical to that of the original reference beam, the reconstructed beams from corresponding points in each 'page' emerge at the same angle and so are focused by the lens L2 into the same point in the image plane, i.e. the reconstructed 'pages' are imaged in the same position.

A further advantage of this imaging system results from the recording of each object point as a single spatial frequency in the hologram. Small lateral displacements of the hologram with respect to the read-out beam do not affect the direction at which the reconstructed waves emerge and so do not affect the position of the image points. There is no requirement therefore for extremely accurate registration of the hologram plate. However the read-out beam must not be misaligned sufficiently to interfere with adjacent holograms.

3. Capacity of the Holographic Memory

The total number of bits which can be stored in the holographic memory is dependent on the number of positions to which the laser beam can be deflected, which controls the number of 'page' holograms which can be stored, and the number of bits stored in each 'page'.

The number of positions depends on the performance of laser beam deflectors using the electro-optic or acousto-optic effects. The beam deflector must give the maximum number of resolved spot positions consistent with the requirements for a switching time of the order of a microsecond and a large deflection efficiency, preferably larger than 10%. A switching time of this order is necessary to make the random access time competitive with that of other memory devices.

The author considers that acousto-optic deflectors offer the best solution, at any rate in the near future; deflectors of this type have been built[2] which produce over 100 resolved spot positions in each direction using a water cell as the active medium, although the switching time is considerably longer than a microsecond. It seems that recent developments in acoustic cell materials[3] will enable switching times to be reduced to 1 μs while maintaining the number of deflection positions.

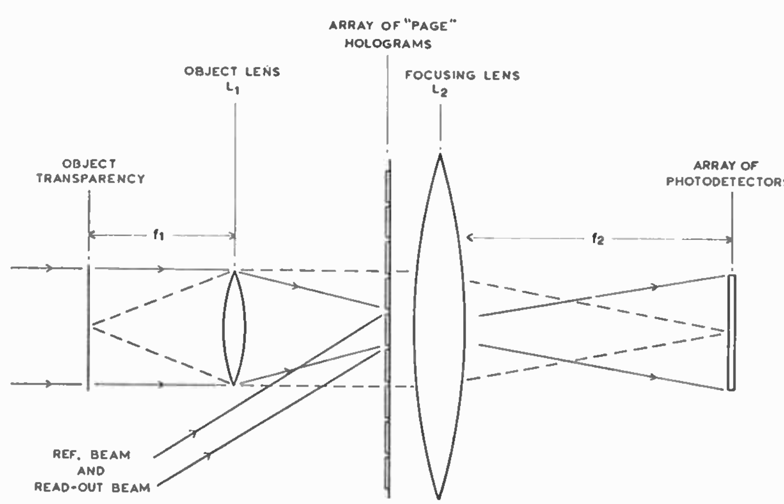


Fig. 2. Optical system for recording and read-out of 'page' holograms.

Taking 100 resolved spot positions as a reasonable target value, the number of beam addresses in a square array can be 10^4.

The number of bits which can be stored in each page hologram is ultimately limited by the light flux available in the reconstruction, which is controlled by the output power of the laser, and the detector sensitivity. Despite their low sensitivity, it is intended to use semiconductor photodiodes as detectors because of the low cost of large integrated arrays.

It has been calculated[14] that a light power of approximately $0{\cdot}3\ \mu W$ is required to produce a detectable output from a silicon photodiode at a wavelength of $0{\cdot}6\ \mu m$ with an integration time of $1\ \mu s$. If the overall efficiency of the optical system including the hologram is 10% (which assumes the use of a phase hologram), then the use of a helium-neon laser with an output power of about 30 mW limits the number of photodetectors to 10^4. On this basis a memory consisting of 10^4 pages and containing 10^4 bits can be constructed using presently available techniques.

Using these figures some of the optical design characteristics of the memory have been investigated and an experimental determination of the quality of page hologram reconstruction has been made.

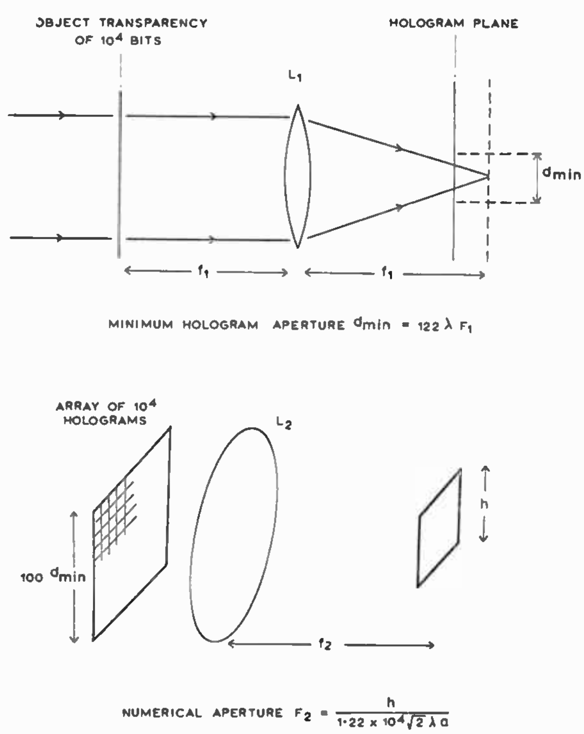


Fig. 3. Dimensions of optical components.

4. Size of Optical Elements Required

In the experimental system used for making holograms illustrated in Fig. 3 the object consists of an array of 10^4 points arranged in a square matrix of 100×100. By making use of the Rayleigh criterion one can show that the minimum diameter of a circular aperture placed in the hologram plane which enables the reconstructed spots to be just resolved is

$$d_{min} = 122\lambda F_1$$

where F_1 is the effective F-number of the object lens, defined as the focal length divided by the width of the object. The value of F_1 used in the experiment to be described was about 4, so the minimum hologram aperture is 0·3 mm for $\lambda = 0{\cdot}6\ \mu m$. This means that theoretically the minimum hologram diameter needed to store each 10^4 bit page is considerably less than 1 mm square.

In practice one expects the required hologram diameter to be somewhat larger than this because the resolution will be affected by aberrations introduced by the lenses and by the hologram itself, also the spot resolution implied by the Rayleigh criterion may not give an adequate 'signal-to-noise' ratio. Thus the size of an array containing 10^8 bits would be about 10 cm square.

One optical component for which the requirements are rather stringent is the focusing lens L2. This lens has to have a diameter large enough to cover the whole information plate and it is also desirable for its focal length to be short so that the size of the 'page' reconstruction is not excessively large. The required F-number for this lens defined as the focal length divided by the lens diameter is therefore rather small.

F_2 can be evaluated by inserting the expression for the reconstruction width h:

$$h = (f_2/f_1)l$$

where l is the width of the original object, into the expression for the size of the hologram array.

F_2 is then found to be:

$$F_2 = \frac{h}{1{\cdot}22 \times 10^4\sqrt{2}\,\lambda a}$$

where the factor a is the ratio of the size of the actual 'page' hologram to the theoretical minimum. We see that F_2 is dependent principally on the size of 'page' reconstruction required. It is quite independent of the size of hologram provided that a is not a function of hologram size.

The size of the reconstruction depends on the size of the available photodetector array. Since an array of 10^4 photodetectors is envisaged it would be convenient if these could be made as an integrated array of silicon photodiodes, preferably with the addressing circuitry included on the same silicon slice using microelectronic techniques. Assuming that the whole array could be made on a single 2·5 cm. (1 in) diameter silicon slice the necessary size of the hologram reconstruction projected on to it would have to be about 2 cm square. The corresponding value of F_2 would then be

$$F_2 = 2/a$$

for $\lambda = 0{\cdot}6\ \mu\text{m}$. Taking the numerical factor a as about 2 the necessary F-number becomes unity.

A lens of this aperture with a diameter of about 10 cm and having low aberrations would be a difficult component to obtain, since large aperture camera lenses have diameters of only a few centimetres.

The problem would be eased if the lens diameter could be reduced. This could be done by reducing the size of the hologram plate and using a larger density of stored information in the hologram. Even so it may be necessary to use a larger photodetector array.

5. Experiment

It was necessary initially to find out whether the reconstruction from very small 'page' holograms could have sufficiently good definition for use in this application, and it was subsequently decided to determine the quality of the reconstruction as a function of the hologram aperture to evaluate the minimum value of a.

The object transparency used in the experiment, shown in Fig. 4, consists of an array of 10^4 transparent squares on a black background; each square is 50 µm wide and the whole array is 1 cm square. Some of the squares were blacked out to simulate 1's and 0's. The transparency was placed in the front focal plane of L1 which was a cine-camera lens of 4 cm focal length.

The hologram plate is placed close to the back focal plane of the lens. This is necessary because the focal plane is the only position in the diffracted field of the object where diffracted light from every point on the object enters the small hologram aperture with approximately equal intensity. Movement of the hologram plate away from the focal plane results in a partial cutting-off of light from the outermost spots in the object with a consequent reduction of recorded intensity. This produces a variation on brightness of the reconstruction as mentioned later.

It is not desirable however to place the hologram plate exactly on the focal plane because the recorded pattern is then the spatial Fourier transform of the object. As a result of the extreme regularity of the distribution of object spots, the spatial Fourier transform contains some spatial frequencies which have much greater intensity than the others. The large dynamic range of intensity is greater than that which can be linearly recorded on the photographic plate.

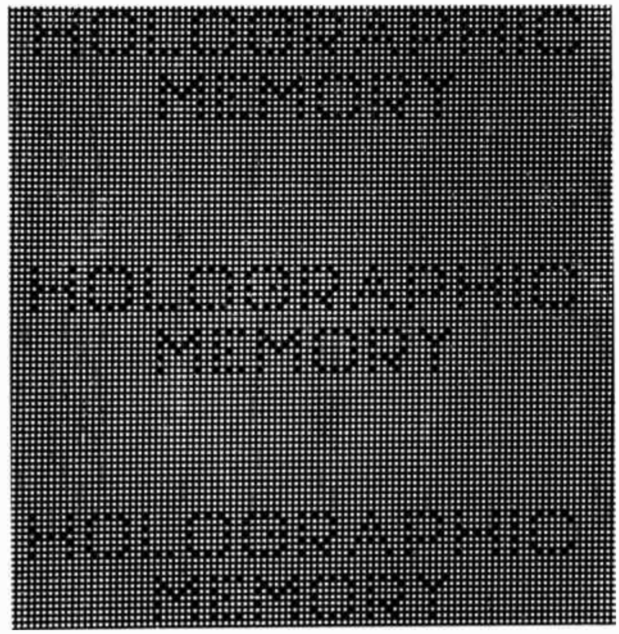


Fig. 4. 10^4-bit object transparency.

Thus it is necessary to make the hologram recording at a position 1 or 2 mm away from the focal plane.

The reference beam, which is a plane wave obtained by tapping-off part of the object illumination using a prism splitter, is incident on the hologram plate at an angle of 40°. The reference beam intensity is made somewhat larger than the object beam intensity to ensure that the modulation recorded on the hologram is on the linear part of the amplitude-transmission transfer characteristic[4] of the plate.

The holographic plates used in the experiments are Agfa-Gevaert type 8E70 high resolution plates for which the normal development procedure is used, although further experiments using bleaching to produce phase holograms are in progress.

Read-out is by a near-parallel Gaussian beam. This is used in preference to a beam of uniform intensity because it is found that any kind of limiting aperture placed in the path of the beam produces some distortion in the reconstruction.†

† The use of a Gaussian beam for hologram read-out implies that the resolution calculation based on the Rayleigh criterion needs modification. This was based on the assumption of the use of point objects and of a read-out beam having uniform intensity distribution across the hologram aperture. A more rigorous calculation which makes allowance for the finite size of object points and for the modification of the hologram transmission function by the Gaussian beam, gives a slightly increased value for the minimum hologram size needed for object points to be just resolved. The theoretical minimum Gaussian beam diameter for spots to be just resolved is about 0·35 mm as against a minimum hologram aperture of 0·3 mm from the simpler theory.

6. Results

The diameter of the Gaussian beam was varied, by inserting lenses of different focal lengths, over a range from 0·3 mm to about 1 mm to simulate variations in hologram aperture. The diameter is defined conventionally as the distance between the $1/e^2$ intensity points. The resulting reconstructions for three diameters of the illuminating beams are shown in Fig. 5; microphotographs showing the structure of individual spots are also shown.

For a beam diameter of 1·1 mm the spots are well resolved and on this scale the reconstruction has an appearance almost indistinguishable from the original object. The spatial noise in the reconstruction, that is, the random fluctuation or intensity from point to point introduced by the holographic process, does not seem to be very large. In the case of the second largest beam diameter of 0·6 mm it is apparent that the size of individual spots has increased but the spots are still adequately resolved. Judging from the

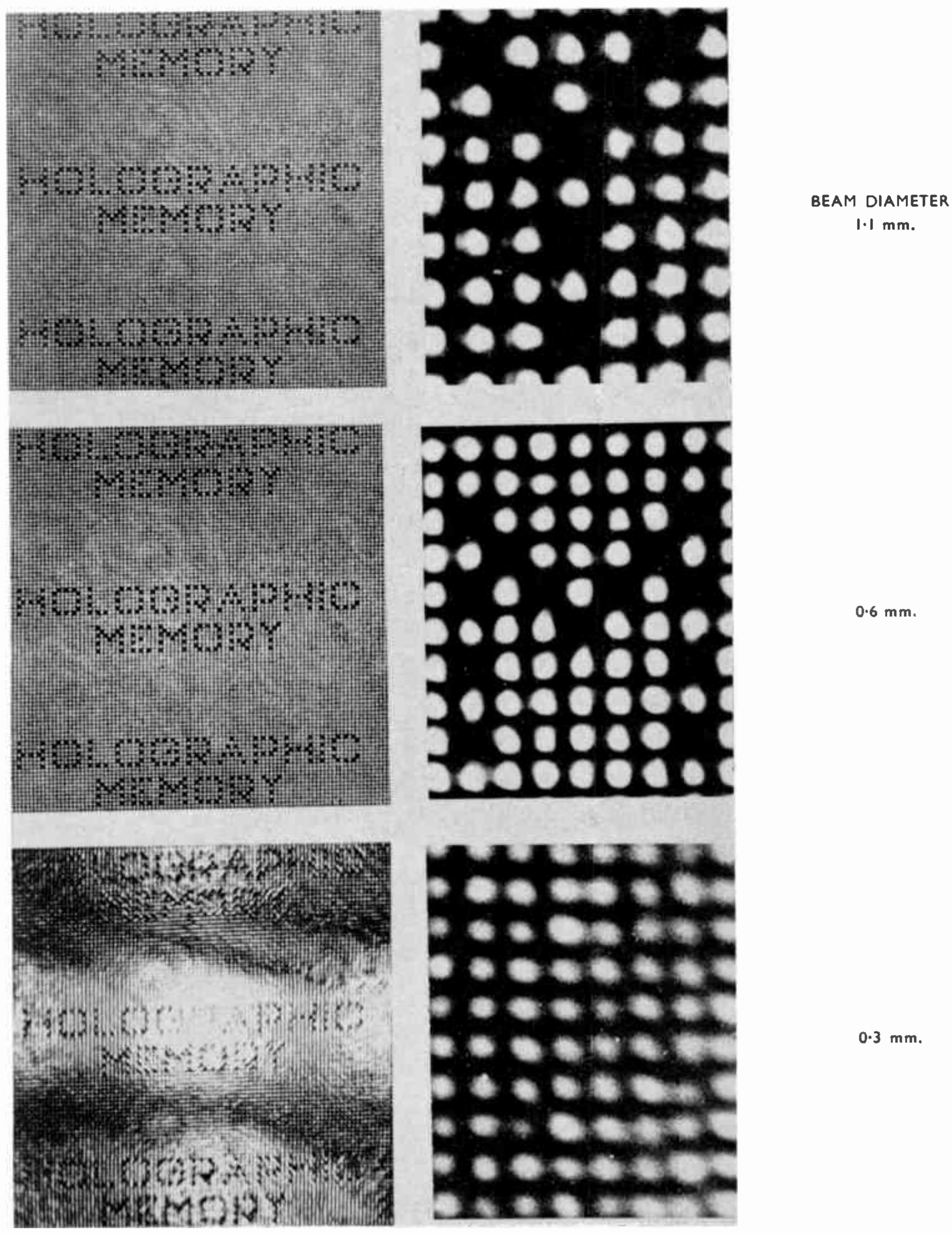


Fig. 5. Reconstruction of transparency using three read-out beam diameters.

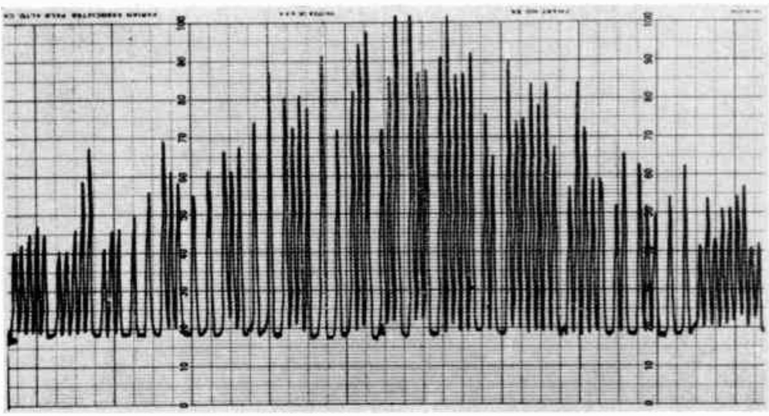


Fig. 6. Intensity distribution across reconstruction from 0·6 mm beam.

photograph the spatial noise is larger than in the previous case but is not large enough for the digital 1's and 0's to be confused.

For a beam diameter of 0·3 mm the individual spots are certainly not well resolved over certain portions of the pattern, also the spatial noise in the reconstruction is very large. Since it is impossible in certain cases to distinguish between 1's and 0's, this beam diameter is below the minimum required for an adequate reconstruction.

To get a more quantitative picture of the spatial noise in the reconstruction intensity measurements were made on individual spots by using a photomultiplier with a small pinhole scanning across the reconstruction plane on a motorized table. The photomultiplier output was plotted using a chart recorder.

The result of tracking across a horizontal line of spots in the middle of the reconstruction from the 0·6 mm beam is shown in Fig. 6. The sampling pinhole in front of the photomultiplier was made sufficiently small to show the structure of individual spots. The chart shows a track across approximately 100 spot positions consisting of digital 1's and 0's, the 1's showing up as spikes and the 0's as gaps in the pattern.

The random fluctuation in intensity from spot is as large as a ratio of 2 : 1 in some places. Superimposed on this is a large scale variation on image intensity which causes the centre to be brighter than the edges. This results from the positioning of the hologram plate during exposure which allows less light to enter the hologram aperture from the outer spots in the object than from the inner ones. In principle this could be corrected for by a corresponding non-uniform illumination of the object transparency when making the hologram.

The random spatial noise in the reconstruction is not entirely contributed by the holographic process as is indicated by the lower chart in Fig. 7. This shows the result of a track across the direct image of the object, without using the holographic step. It is clear that appreciable noise was present in the object beam used to make the hologram, but it was found difficult to remove this.

Although the noise in the reconstruction is not large enough to affect the discrimination between 1's and 0's in a threshold circuit it would be necessary in this

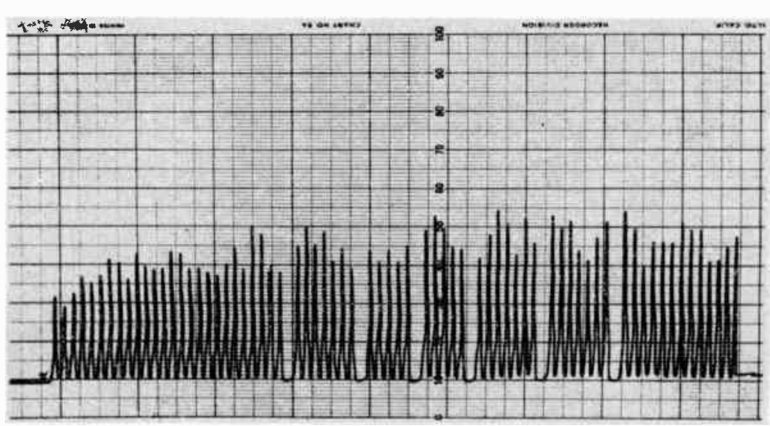


Fig. 7. Intensity distribution across direct image of object.

case to set the threshold circuit at about a factor of four below the peak intensity in the reconstruction, to get satisfactory discrimination.

It is clear that the minimum hologram size which produces a reconstruction of adequate quality is about 0·6 mm in these experiments. This is about a factor of two larger in linear dimensions than the theoretical minimum size. A complete array of 10^4 of these holograms would be somewhere between 6 and 10 cm wide allowing for a small margin between holograms. This represents a rather large diameter for the focusing lens which could be reduced by further reduction in size of 'page' holograms. It is probable that by using an object lens of shorter focal length, a further reduction of two in the hologram size could be obtained without difficulty.

7. Acknowledgments

The author wishes to thank his colleague Dr. K. Firth for useful discussions throughout the course of this work and for suggesting alterations to the manuscript and the Director of Research of the Marconi Company Limited for permission to publish the paper.

8. References

1. Smits, F. M. and Gallaher, L. E., 'Design considerations for a semipermanent optical memory', *Bell Syst. Tech. J.*, **46**, No. 6, p. 1267, July–August 1967.
2. Korpel A., *et al.*, 'A television display using acoustic deflection and modulation of coherent light', *Applied Optics*, **5**, p. 1667, October 1966.
3. Pinnow, D. A. and Dixon, R. W., 'Alpha-iodic acid: a solution-grown crystal with a high figure of merit for acousto-optic device applications', *Applied Phys. Letters*, **13**, p. 156, 15th August 1968.
4. Anderson, L. K. and McMurtry, B. J., 'High speed photo-detectors', *Applied Optics*, **5**, p. 1573, October 1966.
5. Friesem, A. A., Kozma, A. and Adams, G. F., 'Recording parameters of spatially modulated coherent wavefronts', *Applied Optics*, **6**, p. 851, May 1967.

Manuscript received by the Institution on 3rd February, 1969. (Paper No. 1294/Comp. 125.)

STANDARD FREQUENCY TRANSMISSIONS—November 1969

(Communication from the National Physical Laboratory)

Nov. 1969	Deviation from nominal frequency in parts in 10^{10} (24-hour mean centred on 0300 UT)			Relative phase readings in microseconds N.P.L.—Station (Readings at 1500 UT)		Nov. 1969	Deviation from nominal frequency in parts in 10^{10} (24-hour mean centred on 0300 UT)			Relative phase readings in microseconds N.P.L.—Station (Readings at 1500 UT)	
	GBR 16 kHz	MSF 60 kHz	Droitwich 200 kHz	*GBR 16 kHz	†MSF 60 kHz		GBR 16 kHz	MSF 60 kHz	Droitwich 200 kHz	*GBR 16 kHz	†MSF 60 kHz
1	− 300·0	− 0·1	+ 0·1	589	502·3	**17**	− 300·1	− 0·1	0	619	520·3
2	− 300·0	—	+ 0·1	589	—	**18**	− 300·0	− 0·1	0	619	521·3
3	− 300·0	—	+ 0·1	589	—	**19**	− 300·1	− 0·1	0	620	522·3
4	− 300·0	− 0·1	0	589	503·1	**20**	− 300·3	− 0·1	0	623	523·7
5	− 300·1	0	+ 0·1	590	503·4	**21**	− 300·1	− 0·2	0	624	525·3
6	− 300·1	− 0·1	+ 0·1	591	504·4	**22**	− 300·3	− 0·2	0	627	527·3
7	− 300·0	0	+ 0·1	591	505·7	**23**	− 300·3	0	0	630	527·3
8	− 300·2	0	0	593	506·1	**24**	− 300·0	—	0	630	—
9	− 300·1	− 0.1	− 0·1	594	507·9	**25**	− 300·2	—	0	632	—
10	− 300·1	− 0·1	0	595	508·8	**26**	− 300·1	− 0·1	0	631	530·5
11	− 300·0	− 0·1	− 0·1	602	510·3	**27**	− 299·9	− 0·1	0	633	528·1
12	− 300·2	0	− 0·1	598	510·3	**28**	− 300·0	0	0	633	528·3
13	− 300·2	− 0·1	0	600	514·6	**29**	− 300·1	0	0	634	528·5
14	− 300·2	—	0	614	—	**30**	− 299·8	0	0	632	528·7
15	− 300·2	− 0·1	0	616	516·6						
16	− 300·2	− 0·2	0	618	519·6						

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^{11}.

Note: The frequency offset of GBR for 1970 will be -300×10^{-10}

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

† Relative to AT Scale; (AT_{NPL} − Station) = + 468·6 at 1500 UT 31st December 1968.

Contributors to this Issue

G. S. Evans (G.1953; M.1957; F.1969) has a wide experience of the electronics industry, spanning its most formative 20 years. His appointments in the technical and commercial spheres of the industry covered d.c. to microwave and high vacuum to solid state. During the past 10 years he has been closely involved with computer-controlled real-time systems and before establishing Warren Point Limited earlier this year was senior project manager of Elliott Automation Computer Systems Ltd. He is a member of the Institutions' Computer Group Committee and has published previous papers in the *Journal*.

A. J. Tomkins (G.1966; M.1967) read for his degree in natural sciences at Emmanuel College, Cambridge, and subsequently read mechanical sciences, specializing in electronics. He joined Elliott Bros. Ltd. as a development engineer, initially being engaged upon circuit design and later on logic design and systems engineering for digital computers. After a short period with Mullard Equipment Ltd. working on the design of special-purpose data loggers, he took up an appointment as project leader with Elliott Traffic Automation Ltd. He is now employed by Warren Point Ltd. in their real-time systems division.

L. J. Bental received his B.Sc. degree in electrical engineering and electronics from the Israel Institute of Technology in 1953 and the M.Eng. degree from the University of Sheffield in 1955. He then joined Elliott-Automation, initially being concerned with special purposes control devices and later with the design of general purpose computers; eventually he became chief engineer and later technical manager of the Company's Computing Division. In 1965 he transferred to the applications side, pioneering the use of computers in new fields, notably laboratory automation and computer aided design. At the time of contributing the present paper he was manager of systems engineering, responsible for specification and implementation of computer projects in a wide variety of applications. Mr. Bental is now market development executive, advising management on new markets and computer applications within Marconi-Elliott Computer Systems Ltd.

Dr. R. M. Langdon graduated in 1966 from the Department of Electronics of the University of Southampton. He remained at the University for a further two years, obtaining his Ph.D. degree in 1968 for research into the characteristics and mechanism of a pulsed helium-neon laser. He is now with the applied physics group of the research division of the Marconi Company Limited where he is working at holographic stores.

R. C. Crafer graduated in physics at the University of Oxford in 1966. He was awarded an S.R.C. Research Studentship in the Department of Physics of the University of Essex in 1966 and he has recently been appointed to a research fellowship in the Applied Physics Department at the University of Hull.

Professor A. F. Gibson received his B.Sc. and Ph.D. degrees from the University of Birmingham in 1944 and 1948 respectively. From 1944 until 1963 he was concerned with radar research, latterly specializing in semiconductors and infra-red technology, at the Telecommunications Research Establishment (now the Royal Radar Establishment), Malvern. He left Government Service with the grade of deputy chief scientific officer to take up the chair of physics at the University of Essex; he was in fact the first member of the academic staff to be appointed at this new university.

M. F. Kimmitt received his B.Sc. degree from Trinity College, Dublin, in 1955 and he worked at the Royal Radar Establishment, Malvern, until 1965. He then obtained his present appointment of chief research officer in the Department of Physics of the University of Essex.

SUBJECT INDEX

Papers and major articles are denoted by printing the page numbers in bold type

INDEX OF PERSONS

Names of authors of papers published in the volume are indicated by bold numerals for the page reference.
Authors of papers which are given in abstract form are denoted by A.
Biographical references are denoted by B. Contributors to discussions are indicated by D.

INDEX OF ABSTRACTS

This index classifies under subject headings the abstracts throughout the volume in 'Radio Engineering Overseas'.