Founded 1925

Incorporated
by Royal Charter 1961

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

VOLUME 42 No. 1 JANUARY 1972

THE RADIO AND ELECTRONIC ENGINEER

The Journal of the Institution of Electronic and Radio Engineers

The Presidential Address of

A. A. Dyson, O.B.E., C.Eng., F.I.E.E., F.I.E.R.E., S.M.I.E.E.E.*

Delivered at the Annual General Meeting of the Institution in London on 2nd December, 1971

In 1922, a few wireless enthusiasts got together and formed an association of radio engineers in Europe. To help their association in its infancy they elected as President a man of distinction. He was not involved, unfortunately, either professionally or commercially in the business of radio engineering. At this period it was the custom of the British to use the term 'wireless' instead of 'radio' to describe the art of this particular branch of technology, 'radio' being the North American term. Perhaps as a result of this, the Radio Association had a very short life and although the original founders paid the penalty of all pioneers who think ahead of industrial capability to exploit new thinking and scientific discovery, it was not their fault or the fault of the President of the Association that we, in this country, did not understand the title of their Association.

At all events, Britain's ability to compromise led from the original Radio Association to the formation of our own Institution in 1925. Compromise in two senses! First, the term 'wireless' was used to describe the Institution; secondly, the Office of President was not created again for seven years!

Thus in 1932—just about 40 years ago—the first President of this Institution took office. Instead of being complacent, Professor F. T. Fawcett suggested that the real field of the Institution would be better defined as Electronic Engineering. This apt description of our future work did not, however, find wide favour in Europe until after the last War. Meantime, Fawcett's seed of an idea had been germinated by our cousins in the USA, who reaped the benefit of applying radio principles to techniques other than audio or visual communication. In this way an adjective was introduced into world language although derived from the description given in 1891 to the phenomena of physics.

Fawcett was a physicist. Before embracing academic life, he was articled to one of the oldest firms in the country and early came into contact with research on the valve—long before wireless telegraphy demonstrated its importance. He then circumnavigated the world and put in some pioneer work on cables.

Fawcett preceded a succession of Presidents who were not only men of distinction but very, very much involved in the business of radio and electronics which, in the engineering sense, remains the sole concern of our Institution.

In the parlance of 1972, I succeed twenty other Chartered Engineers who have been Presidents of this Institution. Every one of them has contributed to the lustre and international recognition of our Institution.

As a junior engineer, I very much valued the opportunity to join a Chartered Institution. The industry then, as now, was highly competitive and all those involved in the manufacture of components and equipment for radio and what would now be called electronic engineering, spent a great part of their time on purely commercial matters. For a young engineer this was a training that has stood me in good stead. This highly competitive spirit that has always pervaded this industry has meant that many fellow members of the Institution have been in commercial competition with me for very many years.

It was, therefore, especially gratifying to me to be elected unopposed as the President of our Institution just twenty-five years after my election as a Member. For this signal honour I thank all Corporate Members and the Institution's Council who so kindly proposed my election.

I am, however, sorry that our Institution persists in having a Presidential Address immediately the holder takes office! I favour such an address at the middle or end of the Presidential term of office—something like an address on 'The State of the Union'. However, in present conditions, I can only call on my personal experience.

Exploitation of Ideas

Firstly, I should like to draw your attention to one area that has always disappointed me in this electronic industry of ours in this country. It is also perhaps true of other industries but I must confine myself to the areas that I know best.

Britain originally had the benefit of Clerk Maxwell's work. A little later Marconi settled here and continued

* Chairman, Erie Electronics Ltd., Great Yarmouth, Norfolk.

with all his major development work. We, as a nation, were too slow in developing these ideas throughout the whole of the public and private sector, both here and abroad, and even at this time were rapidly overtaken by other countries.

I found, as a young engineer, similar developments in the field of loudspeakers, valves and amplifiers where this country developed to a production stage very viable ideas but were then far too slow to exploit their full manufacturing potential. It may even be argued that in both radar and television in both of which we have been pioneers, we are not the world leaders that we should be. I find this particularly disappointing, for electronics is a prime example of the sort of industry in which this country should be leading the world. Its material content is not so high that we have to pay the penalty of a heavy import of material. The main investment required is in properly trained young men and women who are capable of maximizing the inventions and developments. We not only need people trained in development engineering—we most of all need people trained in the management of the enterprises who are able to manufacture and sell world-wide such developments.

Technical Education

I have always been very interested in training for this industry. Apart from the Defence Services, there was little organized education in radio and electronics until after the War; then, following the lead of the University of Southampton, a number of Universities eventually established special Departments in Communications, Radio or Electronics. In addition, of course, very many new Universities have been established; indeed, it seems that one of the status symbols of County life is to have or advocate the setting up of a University.

Thus, we now have a state where, in general, a University education is open to almost everyone. Have we not, however, developed a system that is just as wasteful as the old system? The products of the Universities, as now shown by our unemployment figures, cannot be absorbed into Industry which, in order to survive, has to innovate and cheapen production. In short, are we wrongly advising, and providing wrong courses for future generations?

One of the engineer's main problems is getting close enough to Universities and colleges to keep them abreast of the very rapid changes that are occurring in industry. This not only concerns technical development which can make obsolete a well-established process overnight, but can also change the skilled manpower needs of the future.

I am hopeful that our studies will concentrate on these areas as a prime objective for uniting the engineering profession.

I have been watching with interest the movements by the professional Institutions within the C.E.I. on setting standards for professional engineers. It is certainly more than reasonable for all the professional Institutions to be working on the same guide-lines. But I am sure that as yet we have not completely solved the problem of the first-rate engineer who has become so by practice and experience, but whose academic qualifications do not meet the requirements of our Institutions. I say this because I am very much aware that a large part of industry is directed and managed by first-rate engineers whose academic background is not sufficient to meet the present requirements for Institution Membership. I also believe that there are now, and will continue to be, non-qualified engineers whose value to this industry will be very great but who, under present circumstances, will have great difficulty in achieving that most honoured cachet of 'Chartered Engineer'.

It may be old-fashioned to talk of apprenticeships, but at all levels (craftsman, technician and engineer) there is considerable personal benefit in learning 'how the job is done' in actual working conditions.

These opportunities must be given to future generations if they are to possess the ability to meet the problems of their generation. Ability to do these things and be capable of anticipating and/or using innovation requires practical as well as theoretical training, and the necessary facilities must be made available if we are to avoid criminal waste of human endeavour and above all, to avoid future unemployment.

Perhaps part of the problems to which I have been referring in University teaching, practical experience and Institution qualification, could, in the long term, be solved by the further development of Technical Universities.

I have been dealing with our friends in America for 40 years and have, therefore, had more than a good opportunity to evaluate that which is good and bad in their various systems and customs. I have always been deeply impressed with the graduates from their Technical Universities. I have also noticed that many of these graduates continue their training at such Universities for at least a further year after their first degree in an associated college on the same campus. This further year concentrates on the study of Business Administration. At the end of this training they produce a graduate engineer not only well-versed in his science, but with an understanding of the basic precepts in the operation of a business.

I feel sure we can benefit from ideas such as these, particularly when we see in America the continuous movement that takes place between Universities and industry. This is stimulated by professional men of high standard in industry being enabled to spend a year in Universities and thus establish a very desirable connexion between such colleges and industry.

Other Institutions, under the collective banner of C.E.I., have given much thought to education and training. In forming the image of the engineer on an absolute base of a university education, however, C.E.I. may in fact be creating not an image but a mirage! For too long professional engineers have had false illusions about the role of the technician. Let us not, in our free society, make the mistake of supposing that people can be filed into categories at quite an early stage in their lives; nor that changing patterns of technical education will result in all Polytechnics and similar colleges becoming incapable of producing competent engineers. In the past, technical colleges have often been the first to appreciate new technologies as many of the older members of this Institution can testify. Freedom starts by giving every individual the opportunity to maximize his talents at any stage of his life.

These are some thoughts on the future training of those people who will be running our industry in the years to come. This is an urgent subject for our industry and country and, therefore, our Institution.

Mergers of Institutions

I now turn to the Institution itself.

At the present moment there are 15 professional Institutions and if the Russian satellite presently on Mars were able to return with a highly intelligent Martian, his observations on the number of Institutions that go to make up the C.E.I. would I fear, be more than rude. They certainly would be if he had not had an opportunity to study the foundation and subsequent history of each Institution.

All the Institutions were obviously started because there was great need for them. The fact that they have grown and flourished indicates the existence of that need which still prevails but is not co-ordinated. To-day there are many areas where individual Institutions overlap one into the other and the genuine, honest query arises as to why we have one or two or three operating in the same field. Our histories tell us why this is so, but the real question is—do we make any changes in the future?

Any change, firstly, must benefit the members. There may be some obvious economic benefits in the operation of joint Institutions and I regret that in some cases such aspects have often far too much weight attached to them and not enough weight attached to the present needs which the individual Institution is presently supplying, and the future capacity to help the younger graduate engineer.

For example, to be of benefit to all members, I believe that the merging of an Institution with another must give the individual member a chance to have a greater say in the running of his profession. This means that an Institution must be completely authoritative. It must be in the forefront of the public's eye and it must also move rapidly on technical matters as well as on any matters affecting the professional life of its members.

Electronics in the Next Decade

I obviously cannot finish without my look into the future. Without such a glance into the crystal ball, no Presidential Address would be complete. I do not wish to baffle members with the detailed changes that may occur, except to say that electronic apparatus is playing an ever-increasing part of our social life. *But* it will be either a much greater annoyance, or benefit, in the *eighties*!

It is said that in a modern military aircraft, a third of the cost of such an aeroplane is allocated to the engines; a further third is the cost of the airframe itself; the remaining third is electronic equipment. I am sure that the amount, complexity and cost of this electronic equipment is little understood by the majority of people.

Electronics plays a great part in every modern industry. We may think to-day that the only electronic part of a car is a radio or cassette. By 1975 there will be, in general production use, electronic systems using hybrid micro-electronics to control ignition, carburation, braking and instrumentation.

In the television field in which I have been interested for so long, we are already seeing the impact of colour television, a process that will accelerate for the next few years. We will be seeing in the internal design of this complicated electronic device, a major change in its method of manufacture. The industry has long sought after a method by which the circuitry may be assembled in a building-block fashion, not only to assist the assembly problems but, of more importance, to enable servicing to be carried out more quickly and more efficiently. To date, economics have always ruled out sub-assemblies in a modular form. We now have technologies such as the silicon integrated circuit and the hybrid thick film integrated circuit, which certainly provide the performance with the ease of assembly. We are now at a point where their economics enable them to be used by this highly-competitive consumer industry. We are also on the threshold of true global television where it may be possible for the individual to directly receive a satellite transmission instead of relying on his local broadcasting station. The political aspects of this are obviously frightening and, once again, we have an example of where technology has raced far ahead of our political ideas.

Our telephone system is still virtually working under the guidance of Alexander Graham Bell and the tremendous expansion that is occurring is mainly concentrated in attempting to satisfy customer demand. When there are signs that at least the supply is catching up with the demand, we may expect to see the development of the wire-less or grid communication system. Indeed, the area of supplying to the home information in printed form using the fundamental video transmission links with the art of electrostatic reproduction, is little heard of outside the development laboratory but could, in fact, be available within the next ten years. This could radically change the present systems of providing a news service on radio, television and the newspapers. This is another example where the fundamental technology already exists but the question of who controls and runs such a service has not yet been discussed.

It is in such political and economic matters that the engineer should exercise his opinion. Regrettably, as shown by the membership of the House of Commons and elsewhere, the engineer takes too little part in public life.

The forward use of the computer has been described by so many other people that I will not repeat the platitudes, especially as it is a little out of favour at the moment. However, it is of the utmost urgency that we decide this country's future role in the computing field.

The area in which electronics will play a more major part and to great effect, and where it is sorely needed, is in the realms of industrial control. This is a non-glamorous subject and an area where, in the last few years, great changes have occurred. The clanking of contactors is being replaced by thyristors, in the measurement of fine tolerance the micrometer is being superseded by the meter, and the incandescent light by the light-emitting diode. This quiet revolution will help industry in its search for greater productivity.

In two years' time we shall be celebrating a quarter of a century of solid state electronics since the transistor effect was first reported by Bardeen and Brattain in 1948 and the junction transistor was introduced by Shockley in 1949.

At a recent meeting of our Institution, Mr. G. W. A. Dummer expressed the view that future development lies in the field of microelectronics and optoelectronics.

In a similar vein, Professor A. L. Cullen, in his address to a Conference on the Economics of the Electronics Industry, said: 'The most interesting development is the advent of solid-state sources of microwaves', and he suggested that perhaps the most exciting prospect for the future was the combination of antenna and solid state technologies to produce active antenna arrays, in which each element of the antenna has its own individual solid state source. This might well replace the current practice of separating the aerial and source by considerable distance and using waveguides to deliver the power to the array elements.

Conclusions

Such changes will have a profound effect on engineers and public spending. However, I would add that one of the major functions of an Institution is to provide a forum for such thoughts as I have expressed.

I believe our Institution must further the recognition of Electronic Engineering and harness the capabilities of all the people so involved, from craftsman to scientist. Only in this way can we play a full part in developing our art for the economic and social benefit of all people.

Unemployment is partly due to our National failure to anticipate technical development. We have also failed to direct our training towards skills having a future demand. We have also failed to provide the necessary finance to exploit these skills. Are these not problems to which our Institution should contribute to the solutions?

As Chartered Engineers we should remember the words of the President of the Council of Engineering Institutions. At our own Dinner last year, His Royal Highness Prince Philip said:

'Engineering is no longer a matter of providing useful structures and helpful gadgets or simply a matter of creating wealth. Technology is now in effective control of our human environment and as engineering is that vital link between science and its application as technology, engineers are the architects of man's future.'

For over forty years our Institution has worked to obtain recognition of the importance of Electronic Engineering to the future of man.

The achievements of the Institution have made all the greater the honour and responsibility of being its President and I hope to be worthy of this distinction.

Address No. 46

The introduction to the Presidential Address by the retiring President, Mr. Harvey F. Schwarz, C.B.E., and the vote of thanks proposed by Sir Arnold Lindley, are printed on page S.4 in the Report of the Annual General Meeting.

U.D.C. 621-52

Predicting Servomechanism Dynamic Errors from Frequency Response Measurements

Mrs. J. M. BROWN, M.A., F.I.M.A.*

Professor D. R. TOWILL,
M.Sc., C.Eng., M.I.Mech.E., F.I.Prod.E., F.I.E.R.E.*

and

P. A. PAYNE (Graduate)*

SUMMARY

A theory is presented for analytically relating frequency domain criteria to time domain performance criteria. Third-order models are shown to adequately predict dynamic errors in a wide variety of high order physical systems with bandwidths differing by more than 100 to 1. Model coefficients lie in a relatively small region of the coefficient plane, suggesting that check-out spot frequencies and tolerances previously established empirically may be of general use.

Good agreement is observed between predictions based on design data and experimental frequency response results obtained under field conditions on tolerancing an electro-hydraulic servo to satisfy dynamic error constraints on time domain inputs. When used in reverse, the technique predicts system filter properties from time domain measurements. Results are applicable to both manual and automatic test procedures.

The general form of low-order model permits a reasonable approximation to frequency domain tolerances by inspection, reducing costly development and field tests.

* *Dynamic Analysis Group, University of Wales Institute of Science and Technology, Cardiff CFl 3NU.*

List of Symbols

M_p	peak amplitude
ω_p	frequency at which peak amplitude occurs (radians/second)
ω_B	bandwidth (radians/second)
S_p	peak of step response
I_p	impulse response peak
t_{Ip}	time at which impulse response peak occurs (seconds)
R_p	ramp peak error
t_{Rp}	time at which ramp peak error occurs (seconds)
P_p	parabolic peak error
a_i	denominator coefficient of system transfer function
b_i	numerator coefficient of system transfer function
$[x_f]$	$(m+n+1)\times(1)$ vector whose elements are frequency domain parameters
$[x_t]$	$(p\times 1)$ vector whose elements are time domain parameters
$\left[\frac{\partial x_f}{\partial A}\right]$	$(m+n+1)\times(m+n+1)$ matrix of sensitivity functions of frequency parameters with respect to the independent system transfer function coefficients, evaluated at their nominal values
$[\Delta A]$	$(m+n+1)\times 1$ vector whose elements are small changes in the independent transfer function coefficients
$\left[\frac{\partial x_t}{\partial A}\right]$	$(p)\times(m+n+1)$ matrix of time domain sensitivity functions
$[\sigma^2_{x_f}]$	$(m+n+1)\times 1$ vector of variances of frequency domain parameters
$[\sigma^2_{x_t}]$	$(p)\times(1)$ vector of variances of time domain parameters

$$\left.\begin{aligned} \omega_0 &= \sqrt[3]{1/a_3} \\ b &= \omega_0^2 a_2 \\ c &= \omega_0 a_1 \end{aligned}\right\} \text{third-order, zero velocity lag, model coefficients}$$

$\bar{y}$	mean value of any parameter y
$\hat{y}$	estimate of any parameter y

1. Introduction

System identification and production test procedures are often based on frequency response data, but a knowledge of dynamic errors such as peak impulse error, peak step error, and peak ramp error is equally important in assessing the goodness of a system. Time and frequency domain responses are related through the Fourier Transform, so that in the design of filters and of feedback control systems it is not difficult to propose empirical relationships between performance criteria such as peak step error and peak amplitude ratio,[1] and impulse peak error and system bandwidth.[2] These relationships sometimes work well and allow frequency domain techniques to be used to design systems to satisfy time domain requirements.

In comparing alternative methods of testing a complex electrohydraulic servomechanism,[3,4] similar empirical relationships were established between the spread of time and frequency response criteria as estimated from the

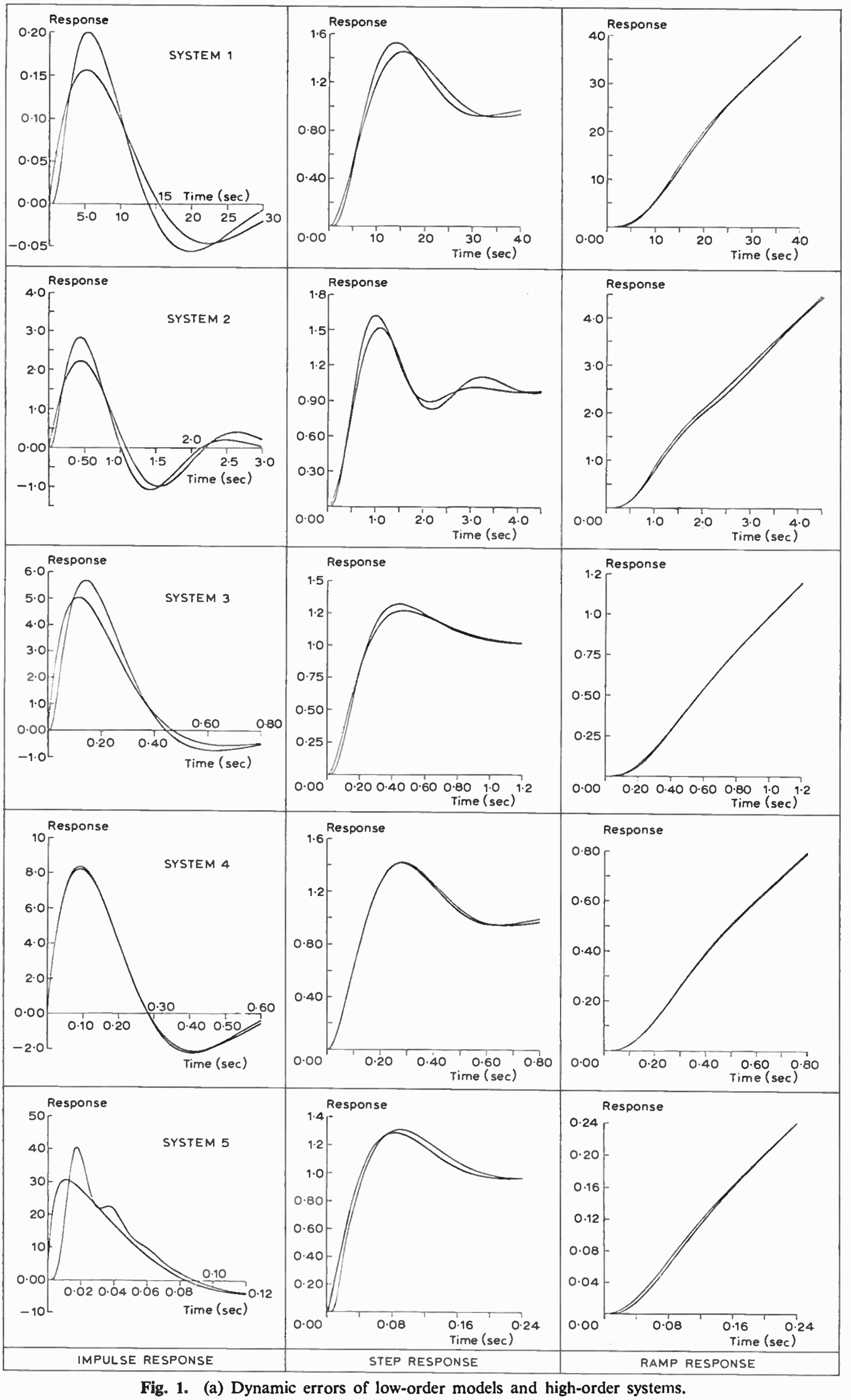
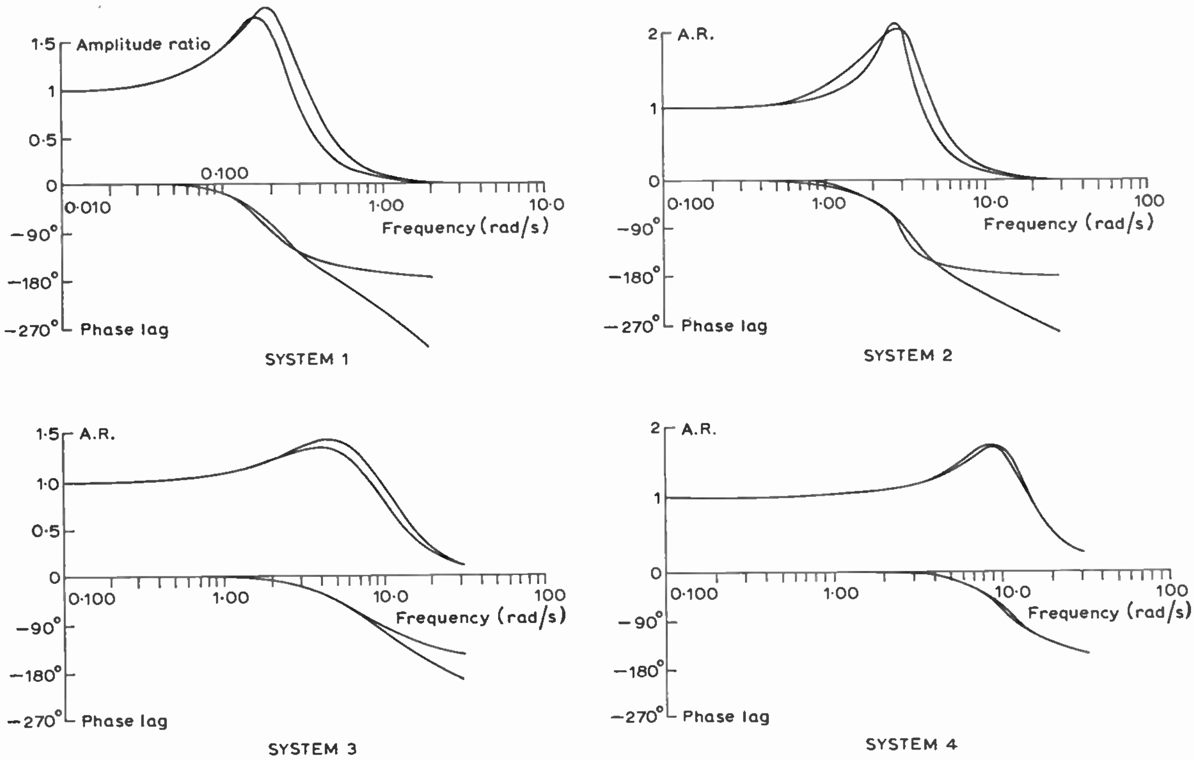


Fig. 1. (a) Dynamic errors of low-order models and high-order systems.



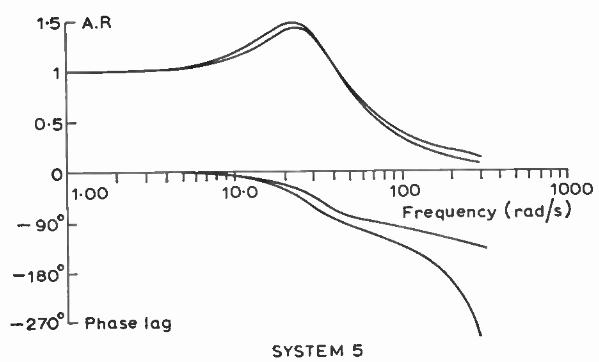


Fig. 1
(b) Frequency response of low-order models and high-order systems.

Table 1. Coefficients of low-order models of the form

$$\frac{1 + c(s/\omega_o)}{1 + c(s/\omega_0) + b\,(s/\omega_0)^2 + (s/\omega_0)^3}$$

derived for a wide range of physical systems

No.	System	Number of system poles and zeros estimated by theoretical analysis	b	c	ω_0	Method of deriving low-order model
1	Aircraft blind landing system[10]	(1, 6)	1·80	1·89	0·182	recurring fraction[13]
2	Aircraft tracking loop[11]	(2, 6)	2·30	2·31	2·30	analogue model[14] fitting to step res ponse
3	Satellite tracker[12]	(1, 5)	2·85	2·86	6·30	updated Bode plot[9]
4	Large gun-position servo[8]	(4, 10)	2·17	1·98	10·7	analogue model[15] in frequency domain
5	Electrohydraulic servo[4]	(1, 8)	3·10	2·50	47·0	least squares curve fitting[16] in frequency domain

production test records for a large batch of the servos. These two papers offered guidance on the number and distribution of measurement points required for go/no-go testing in both time and frequency domains. Furthermore, the expected spread in dynamic performance is predicted from a small sample taken early during the production run, thus indicating a potential quality control application for dynamic measurements. In this paper the results previously obtained are generalized, showing how they can be utilized with confidence not only for developing automatic and manual test procedures but also for specifying tolerances.

2. Steps Required for Generalization of Previous Results

The authors believe that three steps are necessary to establish the generalization discussed in the previous section:

(1) it must be shown that widely differing systems may be adequately described by transfer functions with similar properties,

(2) a mathematical relationship must be established between frequency domain criteria and time domain criteria, and

(3) the techniques must be shown to work under adverse conditions typically met in field testing.

This paper provides these steps via the concept of low-order models.

3. Modelling of Physical Systems

Many physical systems may be adequately modelled using a linear transfer function,

$$\frac{\theta_0}{\theta_i} = \frac{\sum_{j=0}^{m} b_j s^j}{1+\sum_{k=1}^{n} a_k s^k} \qquad \ldots\ldots(1)$$

where the transfer function may vary with the level of the input stimulus. If the transfer function does vary with input level, the model will be appropriate in either the describing function sense or the small perturbation sense.[5]

In this paper the input level considered is that chosen by the user after full-scale tests undertaken at the development stage. A model for predicting dynamic errors need not involve a high-order transfer function. Typically the model chosen must provide a reasonable fit to experimental data in the frequency range between 0·2 and 2 times bandwidth.

There is considerable advantage in choosing the particular low-order model defined by

$$\frac{\theta_0}{\theta_i}(s) = \frac{1+c(s/\omega_0)}{1+c(s/\omega_0)+b(s/\omega_0)^2+(s/\omega_0)^3} \qquad \ldots\ldots(2)$$

i.e. in equation (1), $b_0 = 1$, $a_1 = b_1 = c/\omega_0$, $a_2 = b/\omega^2{}_0$, $a_3 = 1/\omega_0{}^3$, $m = 1$, $n = 3$.

The normalized coefficients b, c and ω_0 allow rapid estimation of model properties via the coefficient plane.[6] Equation (2) may at first sight appear too simple to describe dynamic errors adequately, but subsequently it will be shown to be an adequate predictor for a wide range of physical systems designed by very different techniques. Furthermore, a recent paper[7] suggests that an excessively complex model may have less 'fit' error to specific experimental data, as defined by some arbitrary mathematical criterion, but is often inferior to a simpler model for a more general prediction of response. Practical experience confirms equation (2) is generally adequate, whereas a second-order model is often not.[9]

4. Low-order Models for Physical Systems

Table 1 lists low-order model parameters for five widely varying physical systems covering a bandwidth range in excess of 100 : 1. No attempt has been made to match a particular model derivation technique to a particular system in order to show the technique to best advantage. On the contrary, it is important to show that many practical methods of model building may be used with confidence. The impulse, step, ramp, and frequency responses for the five systems and their low-order models, numbered as in Table 1, are shown in Fig. 1, and confirm the validity of the models. The electrohydraulic servo response shown is modelled from frequency domain data for one servo from a production batch. Since there is a secondary resonance, the third-order model can hardly be expected to predict impulse response oscillations, even though the step response error is well predicted.[17] The gun position servo results are those predicted from high and low-order analogue models established for sinusoidal stimuli superimposed on a slow constant-speed training motion.[8]

5. A Preferred Coefficient Plane Region

Normalized coefficients for the systems of Table 1 are shown in Fig. 2 with results for 30 electrohydraulic servos added to indicate the magnitude of movement in the coefficient plane due to production spread. The

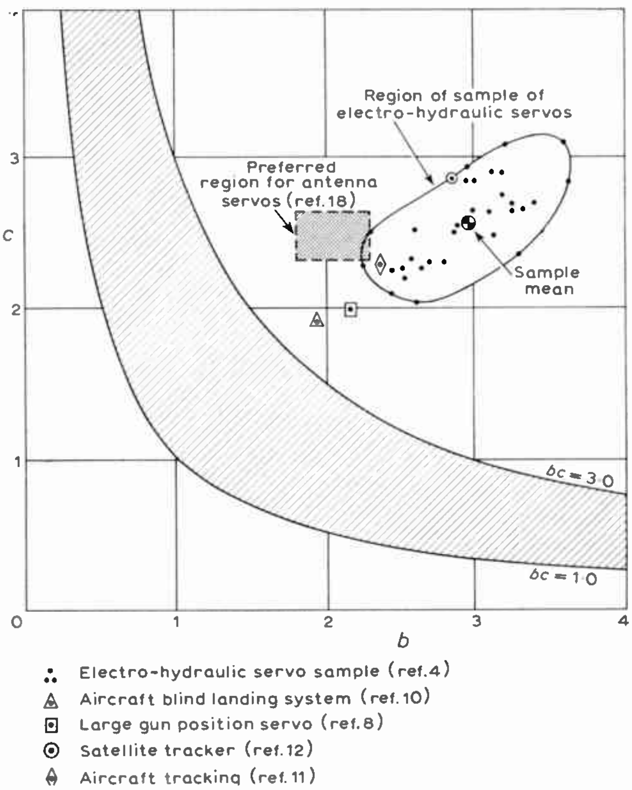


Fig. 2. Low-order model coefficients for various physical systems.

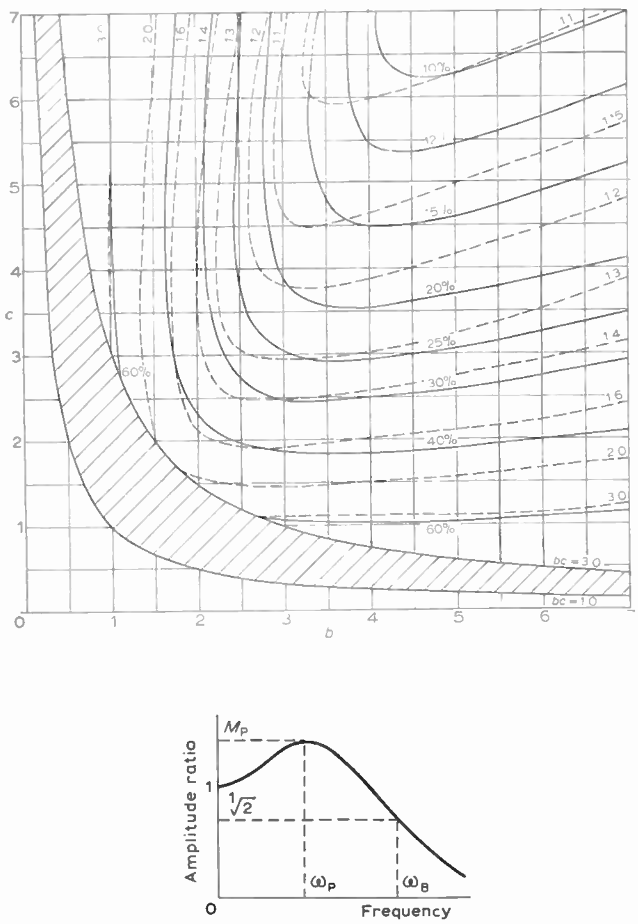


Fig. 3. [Peak amplitude and percentage overshoot contours.
---- M_p —— P.O.

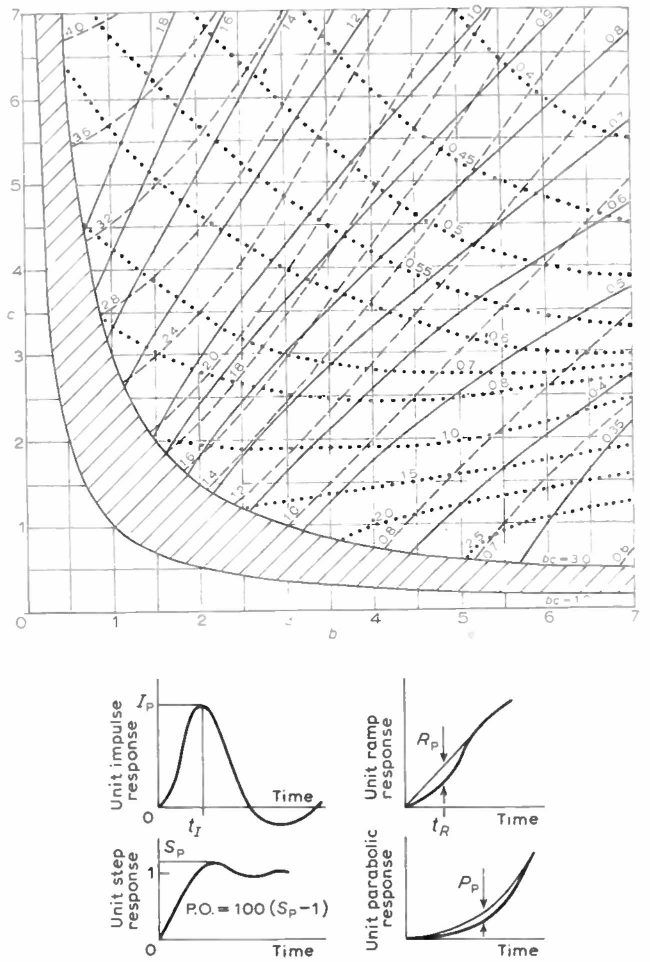


Fig. 4. Bandwidth, impulse peak and impulse time contours.
— — ω_B/ω_0 —— I_p/ω_0 .... $\omega_0 t_{Ip}$

rectangle superimposed represents an area for optimum design of antenna servos in the presence of command, noise, and disturbance inputs.[18] Figure 2 therefore confirms that irrespective of method of system design, there is a strong probability that a low-order model can be found which will lie in a relatively small region of the coefficient plane.

6. Coefficient Plane Contours of Dynamic Performance Criteria

Commonly-used dynamic performance criteria are plotted in Figs. 3 to 8. The quantities peak amplitude ratio (M_p), maximum percentage overshoot (P.O.), the ratio bandwidth divided by peak frequency (ω_B/ω_p), and the product of impulse peak error and ramp peak error ($I_p R_p$), are all independent of ω_0. Response times and ramp peak error (R_p) are inversely proportional to ω_0 whilst peak parabolic error (P_p) is inversely proportional to ω_0^2. Bandwidth, peak frequency, and impulse peak error are proportional to ω_0. By inspection it can be seen that certain time and frequency domain parameters are closely related. For example, in Fig. 4, in the region $2 \leqslant b \leqslant 5$, $1 \leqslant c \leqslant 4$, impulse peak error and bandwidth contours are nearly parallel, thus showing that the results of reference 3 in which the standard deviations of these two quantities are found to be similar over the production run is not coincidental. A formal proof of such relationships is offered in a later section. Numerical calculations, however, will make full use of the graphical information contained in Figs. 3 to 8.

7. Estimation of Model Coefficients from Frequency Domain Data

If peak frequency, bandwidth and peak amplitude ratio are available from frequency domain data, the contours of Fig. 7 may be used to estimate b and c, and Fig. 5 or 6 subsequently used to evaluate ω_0. The model thus obtained will certainly be exact at the two frequency points chosen, and should give a reasonable curve fit over a frequency range up to twice bandwidth. Sometimes the contours of Fig. 7 will have two intersections giving two distinct combinations of b and c, but the two points will either be so close together as to have negligible effect on the time domain parameters, or, if well separated, a high degree of probability can be attached to the point closest to the design value. Even for the intersection point well away from the design value, a good prediction of time domain parameters is possible, as witnessed in Table 2, which is the result for one particular electrohydraulic servo.

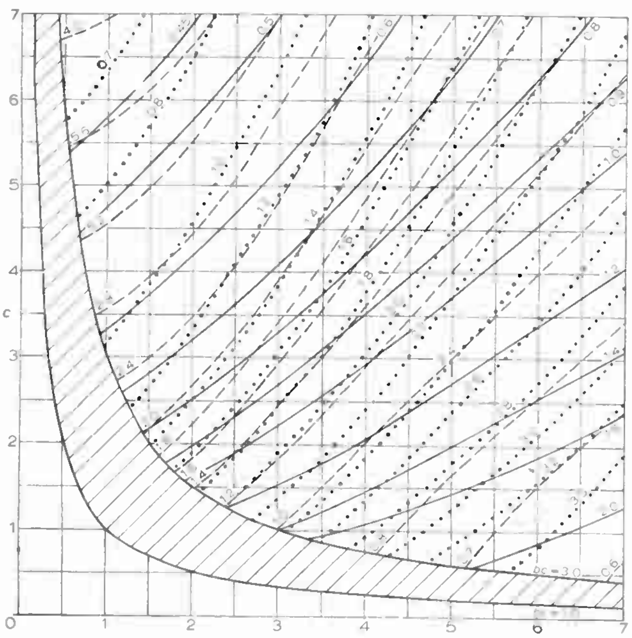


Fig. 5. Bandwidth, ramp peak and ramp peak time contours.
—·— ω_B/ω_0 —— $\omega_0 R_p$ .... $\omega_0 t_R$

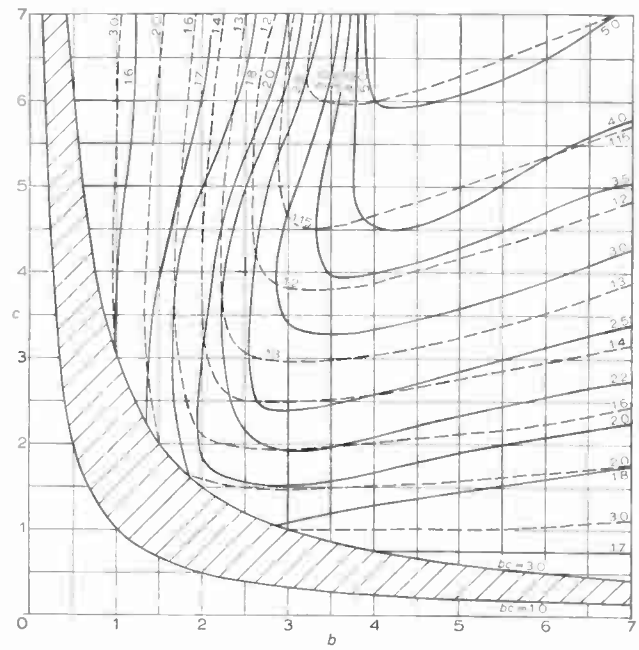


Fig. 7. Peak amplitude and bandwidth/peak frequency contours.
—— M_p —— ω_B/ω_P

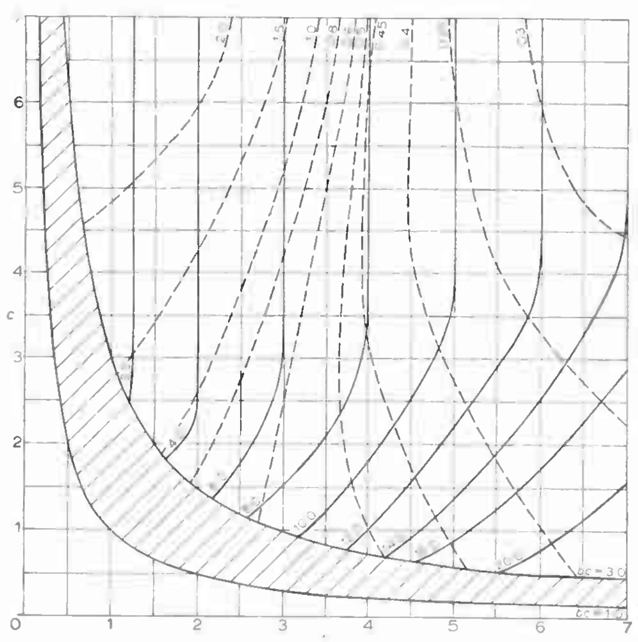


Fig. 6. Peak frequency and maximum parabolic error contours.
---- ω_p/ω_0 —— α_0^2/P_p

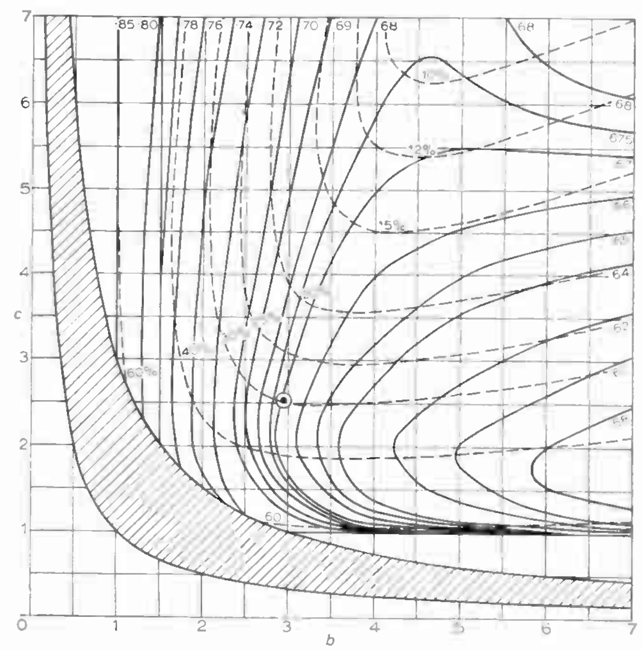


Fig. 8. Ramp peak error × impulse peak and percentage overshoot contours.
—— $I_p R_p$ —— P.O.
⊙ Denotes (b, c) for example of Appendix 2

Table 2. Estimation of model coefficients from coefficient plane contours

(From data $M_p = 1{\cdot}40$, $\omega_p = 26{\cdot}4$ rad/s, $\omega_B = 65{\cdot}8$ rad/s)

	Combination A	Combination B	True value
b	2·60	4·75	2·60
c	2·52	2·70	2·52
ω_0	40·0	65·8	40·0
I_p	31·6	35·5	31·4
S_p	0·31	0·28	0·31
R_p	0·023	0·018	0·022

8. Relating Time and Frequency Domain Performance Criteria

If

$$[\Delta x_f] \triangleq \begin{bmatrix} \Delta x_{f1} \\ \Delta x_{f2} \\ \vdots \\ \Delta x_{f(m+n+1)} \end{bmatrix}$$

represents a vector of small changes in the frequency domain parameters caused by small changes in the independent system transfer function coefficients Δa_k ($k = 1, \ldots, n$), Δb_j ($j = 0, 1, \ldots, m$), and

$$[\Delta x_t] \triangleq \begin{bmatrix} \Delta x_{t1} \\ \Delta x_{t2} \\ \vdots \\ \Delta x_{tp} \end{bmatrix}$$

be the ($p \times 1$) vector for small changes in time domain parameters, then it is shown in Appendix 1 that the time and frequency domain parameters are related as follows

$$[\Delta x_t] = \left[\frac{\partial x_t}{\partial A}\right]\left[\frac{\partial x_f}{\partial A}\right]^{-1}[\Delta x_f] \qquad \ldots\ldots(3)$$

$\left[\frac{\partial x_t}{\partial A}\right]$ is the sensitivity matrix relating changes in time domain criteria to independent transfer function coefficients, whilst

$\left[\frac{\partial x_f}{\partial A}\right]$ is the sensitivity matrix relating changes in frequency domain criteria to independent transfer function coefficients.

Table 3. Performance criteria derivatives in terms of coefficient plane data

FREQUENCY DOMAIN CRITERIA

Model parameter	Criterion: ∂M_p	$\partial\omega_B$	$\partial\omega_p$
$/\partial a_1$	$\omega_0 \dfrac{\partial M_p}{\partial c}$	$\omega_0^2 \dfrac{\partial(\omega_B/\omega_0)}{\partial c}$	$\omega_0^2 \dfrac{\partial(\omega_p/\omega_0)}{\partial c}$
$/\partial a_2$	$\omega_0^2 \dfrac{\partial M_p}{\partial b}$	$\omega_0^3 \dfrac{\partial(\omega_B/\omega_0)}{\partial b}$	$\omega_0^3 \dfrac{\partial(\omega_p/\omega_0)}{\partial b}$
$/\partial\omega_0$	$\dfrac{1}{\omega_0}\left[2b\dfrac{\partial M_p}{\partial b} + c\dfrac{\partial M_p}{\partial c}\right]$	$\left[\dfrac{\omega_B}{\omega_0}\right] + \left[2b\dfrac{\partial(\omega_B/\omega_0)}{\partial b}\right] + \left[c\dfrac{\partial(\omega_B/\omega_0)}{\partial c}\right]$	$\left[\dfrac{\omega_p}{\omega_0}\right] + \left[2b\dfrac{\partial(\omega_p/\omega_0)}{\partial b}\right] + \left[c\dfrac{\partial(\omega_p/\omega_0)}{\partial c}\right]$

TIME DOMAIN CRITERIA

Model parameter	Criterion: ∂I_p	∂S_p	∂R_p
$/\partial a_1$	$\omega_0^2 \dfrac{\partial(I_p/\omega_0)}{\partial c}$	$\omega_0 \dfrac{\partial S_p}{\partial c}$	$\dfrac{\partial(\omega_0 R_p)}{\partial c}$
$/\partial a_2$	$\omega_0^3 \dfrac{\partial(I_p/\omega_0)}{\partial b}$	$\omega_0^2 \dfrac{\partial S_p}{\partial b}$	$\omega_0 \dfrac{\partial(\omega_0 R_p)}{\partial b}$
$/\partial\omega_0$	$\left[\dfrac{I_p}{\omega_0}\right] + \left[2b\dfrac{\partial(I_p/\omega_0)}{\partial b}\right] + \left[c\dfrac{\partial(I_p/\omega_0)}{\partial c}\right]$	$\dfrac{1}{\omega_0}\left[2b\dfrac{\partial S_p}{\partial b} + c\dfrac{\partial S_p}{\partial c}\right]$	$\dfrac{1}{\omega_0^2}\left[-\omega_0 R_p + 2b\dfrac{\partial(\omega_0 R_p)}{\partial b} + c\partial\dfrac{(\omega_0 R_p)}{\partial c}\right]$

If any of the numerator coefficients are linearly dependent on the denominator coefficients then the order of the matrices involving these coefficients will be reduced accordingly.

Convenient frequency domain performance criteria, particularly at the design stage, are M_p, ω_B and ω_p, whilst time domain performance criteria of considerable importance are I_p, S_p, R_p; hence the method will be illustrated by estimating I_p, S_p and R_p from a knowledge of M_p, ω_B and ω_p. In doing so we wish to use graphical information from the coefficient plane in order to reduce the work involved, and to obtain a physical understanding of cause and effect. Because of the form of model chosen, $b_1 = a_1$ and there are only three independent transfer function coefficients. For this analysis it is preferable to use ω_0 in place of a_3, thus,

$$[\Delta A] = \begin{bmatrix} \Delta a_1 \\ \Delta a_2 \\ \Delta\omega_0 \end{bmatrix}, \quad [\Delta x_f] = \begin{bmatrix} \Delta M_p \\ \Delta\omega_B \\ \Delta\omega_p \end{bmatrix}, \quad [\Delta x_t] = \begin{bmatrix} \Delta I_p \\ \Delta S_p \\ R_p \end{bmatrix}$$

with corresponding sensitivity matrices,

$$\left[\frac{\partial x_t}{\partial A}\right] = \begin{bmatrix} \dfrac{\partial I_p}{\partial a_1} & \dfrac{\partial I_p}{\partial a_2} & \dfrac{\partial I_p}{\partial \omega_0} \\ \dfrac{\partial S_p}{\partial a_1} & \dfrac{\partial S_p}{\partial a_2} & \dfrac{\partial S_p}{\partial \omega_0} \\ \dfrac{\partial R_p}{\partial a_1} & \dfrac{\partial R_p}{\partial a_2} & \dfrac{\partial R_p}{\partial \omega_0} \end{bmatrix}$$

$$\left[\frac{\partial x_f}{\partial A}\right] = \begin{bmatrix} \dfrac{\partial M_p}{\partial a_1} & \dfrac{\partial M_p}{\partial a_2} & \dfrac{\partial M_p}{\partial \omega_0} \\ \dfrac{\partial \omega_B}{\partial a_1} & \dfrac{\partial \omega_B}{\partial a_2} & \dfrac{\partial \omega_B}{\partial \omega_0} \\ \dfrac{\partial \omega_p}{\partial a_1} & \dfrac{\partial \omega_p}{\partial a_2} & \dfrac{\partial \omega_p}{\partial \omega_0} \end{bmatrix}$$

Table 3 lists the individual elements of the sensitivity matrix in terms of coefficient plane derivatives. The derivatives are evaluated about the nominal transfer function as perceived by the designer, or about the mean values obtained for early production tests. For the sample of 30 electrohydraulic servos, estimation of the sensitivity matrices about mean values, and solution of equation (3) leads to the influence coefficients of Table 4. It is immediately obvious from the Table and from Fig. 9 that high correlation exists between S_p and M_p, and between I_p and ω_B, as expected from physical reasoning. Thus

$$\Delta\hat{I}_p = 0{\cdot}520\Delta\omega_B \qquad \text{......(4)}$$

$$\Delta\hat{S}_p = 0{\cdot}475\Delta M_p \qquad \text{......(5)}$$

are satisfactory estimates of ΔI_p and ΔS_p respectively. In equation (4) the term $-0{\cdot}2\Delta M_p$ has been neglected. This is justified because although the coefficients associated with ΔM_p and $\Delta\omega_B$ are of the same order, the absolute values of M_p and ω_B are not. Thus 10% changes in M_p or ω_B will cause $\Delta\hat{I}_p$ to vary by 0·028 and 3·64 respectively. (Reference 3 has already established that percentage variations in M_p and ω_B are roughly equal when estimated over a specific production run.) ΔR_p is not so strongly correlated with any one of the three frequency domain parameters chosen, and to predict ΔR_p with reasonable accuracy it is necessary to use all three frequency domain parameters. From Fig. 9 a good prediction is seen to be

$$\Delta\hat{R}_p = 10^{-3}[-15{\cdot}1\Delta M_p - 0{\cdot}526\Delta\omega_B + 0{\cdot}566\Delta\omega_p] \qquad \text{......(6)}$$

Table 4. Influence coefficients for time domain prediction from frequency domain criteria

(Perturbations about $b = 2{\cdot}9$, $c = 2{\cdot}55$, $\omega_o = 46{\cdot}3$)

	Time domain		
	∂I_p	∂S_p	∂R_p
Frequency domain			
$/\partial M_p$	−0·2	0·475	−0·015
$/\partial \omega_B$	0·52	−0·00065	−0·00053
$/\partial \omega_p$	−0·085	0·0020	0·00057

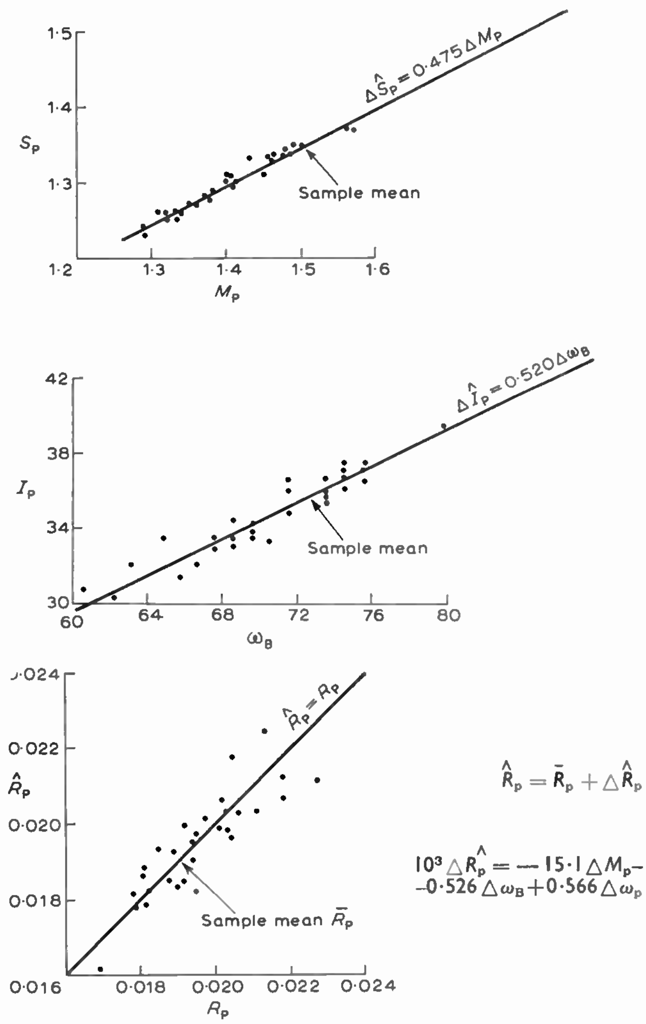


Fig. 9. Estimation of dynamic errors from frequency response for sample of electrohydraulic servos.

The strong correlation between M_p and S_p could have been predicted by inspection of Fig. 3, since in the region of the sample mean the M_p and S_p contours are parallel. Similarly, Fig. 4 shows strong correlation between ω_B and I_p.

9. Estimating Performance Criteria Variance

Variances are also readily estimated by a method similar to that used for correlation of time and frequency domain performance criteria. If

$$[\sigma^2_{x_f}] \triangleq \begin{bmatrix} \sigma^2_{x_{f1}} \\ \vdots \\ \sigma^2_{x_{f(m+n+1)}} \end{bmatrix}, \quad [\sigma^2_{x_t}] \triangleq \begin{bmatrix} \sigma^2_{xt_1} \\ \vdots \\ \sigma^2_{xt_p} \end{bmatrix},$$

As before, the order of $[\sigma^2_{xf}]$ and $[\sigma^2_{x_t}]$ is reduced if any of the bj coefficients are related to any of the a_i coefficients.

Then it is shown in Appendix 1 that the variances are related by

$$\left[\sigma^2_{x_t}\right] = \left[\left(\frac{\partial x_t}{\partial A}\right)^2\right]\left[\left(\frac{\partial x_f}{\partial A}\right)^2\right]^{-1}\left[\sigma x_f^2\right] \quad \ldots\ldots(7)$$

As an application of equation (7), σ_{I_p}, σ_{S_p} and σ_{R_p} are estimated given $\sigma_{M_p} = 0{\cdot}08$, $\sigma_{\omega_B} = 4{\cdot}4$, $\sigma_{\omega p} = 2{\cdot}4$, for the same batch of electrohydraulic servos. Table 5 compares predictions with results obtained by digital computation of the impulse, step, and ramp responses and subsequent processing of the 30 values of R_p, I_p and S_p. Agreement is excellent.

10. Tolerancing the System Frequency Response

An important application of the material in this paper is the guidance given to system testers on tolerancing experimental measurements to constrain dynamic errors, if testing in the frequency domain using a transfer function analyser,[19] and constraining filter characteristics if testing in the time domain using a cross-correlator.[20] The method is based on equation (7) and will be illustrated by tolerancing the system frequency response amplitude ratio at three spot frequencies in order to constrain I_p, S_p and R_p to lie between certain values.

By superimposing coefficient plane contours of constant $I_p R_p$ on those of S_p and using a method similar to Section 7, it may be shown that for each combination of I_p, S_p and R_p, b, c and ω_0 can be estimated with a high degree of confidence. Unfortunately, extreme values of I_p, S_p and R_p do not necessarily occur for the same combination of b, c and ω_0, and even if these do, it does not follow that x_f has an extreme value. This means that the tolerances must be determined on a statistical basis by evaluating standard deviations σ_{x_f} from equation (7) rather than estimating a linear change from equation (3). Care must be taken in choosing realistic values for the constraints as an arbitrary choice might lead to negative values of the variances in a_1, a_2 or ω_0.

Suitable spot frequencies for the electrohydraulic servo[3] are 18, 50 and 100 rad/s. These frequencies are approximately 25%, 75% and 150% of system bandwidth, and the evidence of Table 1 suggests that these *ratios* may well be suitable for a wide range of applications. For a general system model, equation (28) of Appendix 1 would be used to solve this problem, but for the third-order model defined by equation (2), considerable simplification is possible by taking full advantage of the properties of the normalization frequency, ω_0. Hence, on manipulation we have,

$$\begin{bmatrix} \sigma^2_{x_{18}} \\ \sigma^2_{x_{50}} \\ \sigma^2_{x_{100}} \end{bmatrix} = CD^{-1}\begin{bmatrix} \sigma^2_{I_p}/\omega_0^2 \\ \sigma^2_{S_p} \\ \omega_0^2\sigma^2_{R_p} \end{bmatrix}$$

where

$$C = \begin{bmatrix} \left(\frac{1}{\omega_0}\frac{\partial x_{18}}{\partial a_1}\right)^2, & \left(\frac{1}{\omega_0^2}\frac{\partial x_{18}}{\partial a_1}\right)^2, & \left(\omega_0\frac{\partial x_{18}}{\partial \omega_0}\right)^2 \\ \left(\frac{1}{\omega_0}\frac{\partial x_{50}}{\partial a_1}\right)^2, & \left(\frac{1}{\omega_0^2}\frac{\partial x_{50}}{\partial a_2}\right)^2, & \left(\omega_0\frac{\partial x_{50}}{\partial \omega_0}\right)^2 \\ \left(\frac{1}{\omega_0}\frac{\partial x_{100}}{\partial a_1}\right)^2, & \left(\frac{1}{\omega_0^2}\frac{\partial x_{100}}{\partial a_2}\right)^2, & \left(\omega_0\frac{\partial x_{100}}{\partial \omega_0}\right)^2 \end{bmatrix}$$

and

$$D = \begin{bmatrix} \left(\frac{1}{\omega_0^2}\frac{I_p}{\partial a_1}\right)^2, & \left(\frac{1}{\omega_0^3}\frac{\partial I_p}{\partial a_2}\right)^2, & \left(\frac{\partial I_p}{\partial \omega_0}\right)^2 \\ \left(\frac{1}{\omega_0}\frac{\partial S_p}{\partial a_1}\right)^2, & \left(\frac{1}{\omega_0^2}\frac{\partial S_p}{\partial a_2}\right)^2, & \left(\omega_0\frac{\partial S_p}{\partial \omega_0}\right)^2 \\ \left(\frac{\partial R_p}{\partial a_1}\right)^2, & \left(\frac{1}{\omega_0}\frac{\partial R_p}{\partial a_2}\right)^2, & \left(\omega_0^2\frac{\partial R_p}{\partial \omega_0}\right)^2 \end{bmatrix} \quad \ldots\ldots(8)$$

From Table 3 it will be observed that both C and D matrices are dependent only on b and c, since the derivatives for x_f follow the same pattern as M_p. As shown in Appendix 2, for perturbations about the electrohydraulic servo sample mean, we have,

$$\left.\begin{aligned} \sigma^2_{x_{18}} &= \frac{0{\cdot}500}{(46{\cdot}3)^2}\sigma^2_{I_p} + 192\sigma^2_{S_p} - 0{\cdot}918(46{\cdot}3)^2\sigma^2_{R_p} \\ \sigma^2_{x_{50}} &= -\frac{9{\cdot}28}{(46{\cdot}3)^2}\sigma^2_{I_p} - 28{\cdot}7\sigma^2_{S_p} + 24{\cdot}2(46{\cdot}3)^2\sigma^2_{R_p} \\ \sigma^2_{x_{100}} &= -\frac{0{\cdot}108}{(46{\cdot}3)^2}\sigma^2_{I_p} - 1{\cdot}05\sigma^2_{S_p} + 1{\cdot}11(46{\cdot}3)^2\sigma^2_{R_p} \end{aligned}\right\} \quad \ldots\ldots(9)$$

Substituting $\sigma_{I_p} = 2{\cdot}2$, $\sigma_{S_p} = 0{\cdot}040$, $\sigma_{R_p} = 0{\cdot}0012$, from Table 5, into equation (9), the predicted spreads in amplitude ratio at the three spot frequencies listed in Table 6 are readily obtained. These spreads compare very favourably with checkout tolerances determined by digital simulation of a wide range of failure conditions for the particular servo under study.

Table 5. Time domain criteria standard deviations evaluated for sample of 30 electrohydraulic servos

	Predicted by partial derivative method given model parameter variances	Calculated from transient responses
σ_{I_p}	2·2	2·2
σ_{S_p}	0·040	0·040
σ_{R_p}	0·0012	0·0015

Because so many real systems can be modelled in the same region of the coefficient plane, as shown in Fig. 2, it may be argued that the numerical values of equation (9) will not change greatly from one design to another, and that to a reasonable approximation the equation may be used as it stands, using appropriate values of ω_0 and time domain variances. The amplitude ratio tolerances would then apply to test frequencies approximately $0{\cdot}40\omega_0$, $1{\cdot}10\omega_0$, and $2{\cdot}20\omega_0$. Equation (9) therefore provides approximate tolerances by inspection.

Table 6. Frequency domain tolerances for time domain constraints on electrohydraulic servo performance of $\sigma_{\mathrm{Ip}} = 2{\cdot}2$, $\sigma_{\mathrm{Sp}} = 0{\cdot}040$, and $\sigma_{\mathrm{Rp}} = 0{\cdot}0012$

ω (rad/s)	18	50	100
x_f amplitude ratio at ω as averaged over sample of 30	1·294	1·040	0·419
σx_f predicted from partial derivatives	0·0369	0·0878	0·0388
Tolerances on x_f set at $\pm 3\sigma_{x_f}$ level	1·18 to 1·41	0·77 to 1·30	0·30 to 0·54
$\pm 3\sigma$ for sample of 30	1·20 to 1·39	0·80 to 1·29	0·28 to 0·56
Tolerances determined in reference (3) based on spread of sample of 30	1·10 to 1·50	0·70 to 1·30	0·25 to 0·60

11. Practical Modification of Checkout Gates due to Test Conditions

In tolerancing system performance for checkout purposes, mean values of gates used at the test bench may be slightly different from those predicted from design data using the techniques described in this paper. There are a variety of reasons:

(a) Not all design assumptions may be valid, such as neglected time constants.

(b) Low-order model is not an exact fit to the high-order theoretical system at the measured frequencies or time delays.

(c) The manufacturing process may not yield means in accordance with design calculations. However, provided the new marginal systems are satisfactory to the user, it may be advantageous to update the mean so that systems are not unnecessarily failed.

(d) Non-linearities may affect measurements, but to a controlled extent, at the input stimulus level chosen for checkout.

(e) Effects of interface equipment.

(f) Problems of access such as the measurement of a rate loop sitting inside an active position loop.

The effect of item (f) in estimating rate loop dynamics is to provide a feedback path K/s, giving a fourth-order system. For low K, the new transfer function factorizes approximately into a high-pass filter $Ks/(1+Ks)$ in series with a third-order system approximately equal to the rate loop transfer function. In practice, for the electrohydraulic servo, $K \simeq 1$, and it can be seen from Fig. 10 that dynamic errors are not greatly affected by the method of test, although clearly steady state errors are vastly different. Provided these contributions do not shift the coefficients b and c of the low-order model away from the preferred region of the coefficient plane (Fig. 2), there should be little cause to change the checkout gate widths, but simply to adjust the means as experimental evidence becomes available.

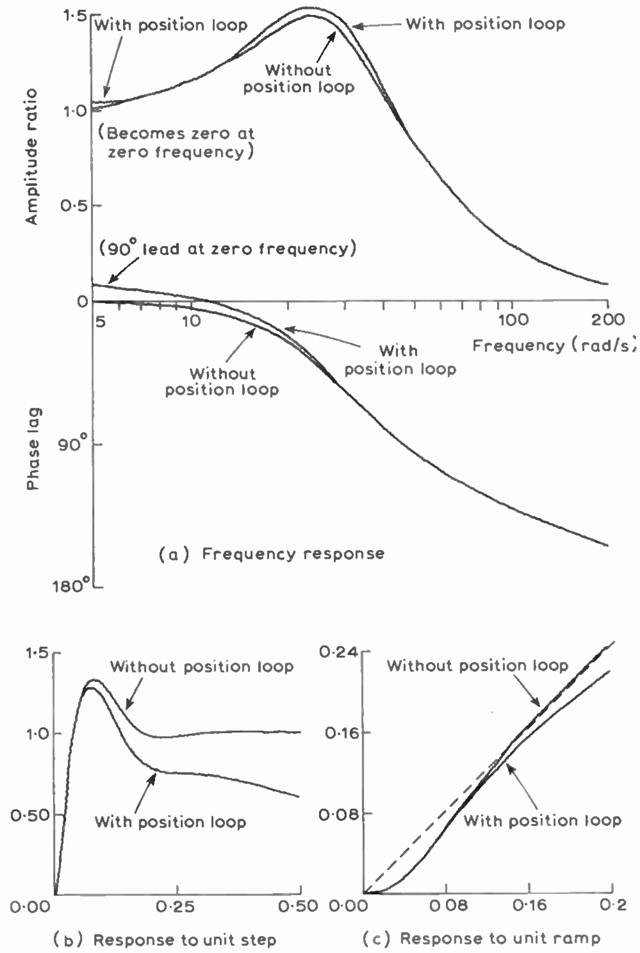
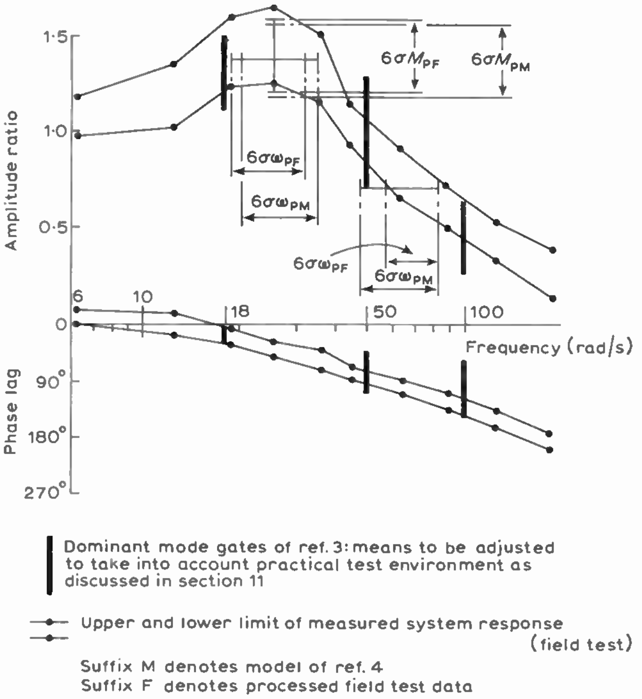


Fig. 10. Effect of position loop on rate loop test.

The degree of adjustment required in the case of the electrohydraulic servo family may be judged by reference to Fig. 11, where checkout gates derived from the low-order model are compared with the results of complete frequency response tests made on 20 members of the family. At all gates except AR_{18} all the tested systems are passed, although changes in gate midpoints are indicated at AR_{100} and Phase $_{100}$. At AR_{18} only one system tested falls outside the low-order model gate, and this would be overcome by an adjustment of the mean value. All the effects (a) to (f) listed above are catered for in the experimental results. 6σ values for M_{p}, ω_{p} and ω_{B} are also compared in Fig. 11 as a final check on prediction, the comparison being between the statistical model of reference 4 and properties of low-order models established from each of the twenty full frequency response tests.



ω_B, M_p and ω_p from design data low order model compared with statistical model and field test mean values

Parameter	Statistical model[4]	Processed field tests	Model from design data[4]
ω_B (rad/s)	70·3	65·5	60·0
M_p	1·40	1·38	1·45
ω_p (rad/s)	27·4	25·2	26·5

Fig. 11. Comparison of field test data (before and after processing) with statistical model and go/no–go gates.

12. Conclusions

Lower-order models adequately predict dynamic errors of high-order systems. Because many practical systems may be modelled by third-order zero velocity lag models, problems such as model identification, time-frequency domain correlation and test measurement tolerancing are greatly simplified using the coefficient plane technique. The generalizations of the paper, justified by application to production test results for a family of electrohydraulic servomechanisms, suggest that dynamic errors may be constrained by spot frequency tests at multiples of about 0·25, 0·75 and 1·50 times system bandwidth, or equivalently at about 0·40, 1·10 and 2·20 times the third-order model normalization frequency. Because so many real systems can be modelled by similar low-order transfer functions, it is suggested that approximate frequency domain tolerances may be obtained by inspecting a simple formula.

Theory also suggests that system filter properties may be inferred by estimating impulse, step, and ramp peak errors, a procedure readily achieved using a cross-correlator and associated circuitry.

13. Acknowledgments

The authors wish to acknowledge the support of the Science Research Council for provision of laboratory facilities, and the Ministry of Defence (Procurement Executive) for experimental facilities, particularly those arranged in industry.

14. References

1. Jaworsky, Z. E., 'Empirical transient formulae', *Electronic Engineering*, **26**, pp. 396–400, September 1954.
2. Biernson, G. A., 'A general technique for approximating transient response from frequency response asymptotes', *Trans. Amer. Inst. Elect. Engrs (Applications and Industry)*, **75**, Part 2, pp. 253-73, 1956.
3. Towill, D. R. and Payne, P. A., 'Frequency domain approach to automatic testing of control systems', Proceedings of Conference on Automatic Test Systems Birmingham, April 1970, pp. 529–48 (I.E.R.E. Conference Proceedings No. 17), *The Radio and Electronic Engineer*, **41**, No. 2, pp. 51–60, February 1971.
4. Payne, P. A., Towill, D. R., and Baker, K. J., 'Predicting servomechanism dynamic performance from limited production test data', *The Radio and Electronic Engineer*, **40**, No. 6, pp. 275–88, December 1970.
5. Towill, D. R., 'Transfer Function Techniques for Control Engineers', p. 430 (Iliffe, London, 1970).
6. Towill, D. R., 'Coefficient plane synthesis of zero velocity lag servomechanisms', *The Radio and Electronic Engineer*, **34**, No. 6, pp. 323-34, December 1967.
7. Taylor, L. W. (Jnr.), 'How complex should a model be?' Joint Automatic Control Conference, Atlanta, Georgia, June 24–26, 1970.
8. Kirby, J. E. D., 'Evaluation of the Transfer Function Measurement Unit' (Final Report), Wayne Kerr Company, Chessington, Surrey, 1968.
9. Towill, D. R. and Mehdi, Z., 'Prediction of the transient response sensitivity of high order linear systems using low order models', *Measurement and Control*, **3**, pp. T1–T9, January 1970.
10. Gill, F. R., 'Integrity of a Civil Blind Landing System with Particular Reference to the Azimuth Channel'. R.A.E. Technical Report No. 65022, 1965.
11. Mehdi, Z., 'Study of the Effects of Component Tolerances on the Performance of Control Systems'. University of Wales, Ph.D. Thesis, February 1971.
12. Wilson, D. R., 'A survey of the developments in antenna servo techniques', *The Radio and Electronic Engineer*, **38**, No. 3, pp. 169–77, September 1969.
13. Chen, C. F. and Shieh, L. S., 'A novel approach to linear model simplification', *Internatl J. Control*, **8**, No. 6, pp. 561–70, 1968.
14. Milsum, J. H., 'Determining transfer functions with analogue computers', *Instrum Soc. Amer. J.*, **7**, No. 4, pp. 54–57, 1960.
15. Levinge, R. W., 'A transfer function computer', *Electronic Engineering*, **37**, No. 446, pp. 218–24, 1965.
16. Payne, P. A., 'An improved technique for transfer function synthesis from frequency response data', *I.E.E.E. Trans. on Automatic Control*, **AC-15**, No. 4, pp. 480–83, August 1970.
17. Towill, D. R. and Payne, P. A., 'Adequacy of low order modelling for a complex electrohydraulic servo', Proceedings Fourth I.E.E.E. Conference on Systems and Circuit Theory, Asilomar, California, November 1970.

18. Nelson, W. L. and Cole, W. J., 'Autotrack control systems for antenna mounts with non-orthogonal axes', *Bell Syst. Tech. J.*, 44, pp. 1367–1403, September 1965.

19. Elsden, C. S. and Ley, A. J., 'A digital transfer function analyser based on pulse rate techniques', *Automatica*, 5, No. 1, pp. 51–60, 1969.

20. Dack, D., 'System identification by on-line correlation', *Control Engineering*, 17, pp. 64–70, March 1970.

21. Ratcliffe, J. F., 'Elements of Mathematical Statistics' (Oxford University Press, 1967).

15. Appendix 1: Derivation of general method of estimating time domain parameters from frequency domain measurements

15.1. *Frequency Domain Definitions*

Suppose that the $(m+n+1)$ coefficients a_k $(k = 1, 2, \ldots, n)$, b_j $(j = 0, 1, 2, \ldots, m)$ of the system transfer function,

$$\frac{\theta_o}{\theta_i}(s) = \frac{\sum_{j=0}^{m} b_j s^j}{1+\sum_{k=1}^{n} a_k s^k} \qquad \ldots\ldots(10)$$

can vary independently.

Define

$$[x_f] \triangleq \begin{bmatrix} x_{f1} \\ x_{f2} \\ \cdot \\ \cdot \\ \cdot \\ x_{f(m+n+1)} \end{bmatrix} \qquad \ldots\ldots(11)$$

as a vector whose elements are an appropriate arbitrary set of measurable frequency domain parameters, such as gain and phase data at specific frequencies.

Let

$$\left[\frac{\partial x_f}{\partial A}\right] \triangleq \begin{bmatrix} \frac{\partial x_{f1}}{\partial a_1}, \frac{\partial x_{f1}}{\partial a_2}, \ldots \frac{\partial x_{f1}}{\partial a_n}, \frac{\partial x_{f1}}{\partial b_0}, \ldots \frac{\partial x_{f1}}{\partial b_m} \\ \frac{\partial x_{f2}}{\partial a_1}, \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial x_{f(m+n+1)}}{\partial a_1}, \ldots\ldots\ldots\ldots\ldots\ldots \frac{\partial x_{f(m+n+1)}}{\partial b_m} \end{bmatrix} \qquad \ldots\ldots(12)$$

be the $(m+n+1)\times(m+n+1)$ matrix of frequency parameter sensitivity functions. Small changes in frequency domain parameters caused by small changes in system parameters can be defined by the vector,

$$[\Delta x_f] \triangleq \begin{bmatrix} \Delta x_{f1} \\ \Delta x_{f2} \\ \cdot \\ \cdot \\ \Delta x_{f(m+n+1)} \end{bmatrix} \qquad \ldots\ldots(\ 3)$$

whilst the changes in system parameters are defined by

$$[\Delta A] = \begin{bmatrix} \Delta a_1 \\ \Delta a_2 \\ \cdot \\ \cdot \\ \cdot \\ \Delta a_n \\ \Delta b_0 \\ \cdot \\ \cdot \\ \cdot \\ \Delta b_m \end{bmatrix} \qquad \ldots\ldots(14)$$

Hence combining equations (12), (13), (14) for first-order changes,

$$[\Delta x_f] = \left[\frac{\partial x_f}{\partial A}\right][\Delta A] \qquad \ldots\ldots(15)$$

If any of the b_j are linear combinations of the a_k so that the number of independent coefficients is q, then $[x_f]$ and $\left[\frac{\partial x_f}{\partial A}\right]$ are of order $(q \times 1)$ and $(q \times q)$ respectively.

15.2. *Time Domain Definitions*

Similarly, let

$$[\Delta x_t] \triangleq \begin{bmatrix} \Delta x_{t1} \\ \Delta x_{t2} \\ \cdot \\ \cdot \\ \cdot \\ \Delta x_{tp} \end{bmatrix} \qquad \ldots\ldots(16)$$

be the $(p \times 1)$ vector of changes in the time domain vector,

$$[x_t] \triangleq \begin{bmatrix} x_{t1} \\ x_{t2} \\ x_{t3} \\ \cdot \\ \cdot \\ \cdot \\ x_{tp} \end{bmatrix} \qquad \ldots\ldots(17)$$

which is required to be estimated.

$$\left[\frac{\partial x_t}{\partial A}\right] \triangleq \begin{bmatrix} \frac{\partial x_{t1}}{\partial a_1} & \cdots & \frac{\partial x_{t1}}{\partial b_m} \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \frac{\partial x_{tp}}{\partial a_1} & \cdots & \frac{\partial x_{tp}}{\partial b_m} \end{bmatrix} \qquad \ldots\ldots(18)$$

is the $(p)\times(m+n+1)$ matrix of time domain parameter sensitivity functions evaluated at their mean values.

Then combining equations (14), (16) and (18) we may write

$$[\Delta x_t] = \left[\frac{\partial x_t}{\partial A}\right][\Delta A] \qquad \ldots\ldots(19)$$

for first-order changes.

15.3 *Time Domain Parameter Estimates*

ΔA may be eliminated between equations (15) and (19) to give the first-order estimate of time domain parameter changes,

$$[\Delta x_t] = \left[\frac{\partial x_t}{\partial A}\right]\left[\frac{\partial x_f}{\partial A}\right]^{-1}[\Delta x_f] \qquad \ldots\ldots(20)$$

Equation (20) is the fundamental equation for estimating first-order variations in time domain parameters from measured variations in appropriate frequency domain parameters chosen so that $\left[\frac{\partial x_f}{\partial A}\right]$ is not a singular matrix.

15.4 *Variance Estimates*

Similar relationships can be derived for the relationship between variances of time and frequency domain parameters.

If

$$[\sigma^2_{x_f}] \triangleq \begin{bmatrix}\sigma^2_{x_{f1}} \\ \vdots \\ \sigma^2_{x_{f(m+n+1)}}\end{bmatrix} \qquad \ldots\ldots(21)$$

and

$$[\sigma^2_{xt}] \triangleq \begin{bmatrix}\sigma^2_{x_{t1}} \\ \vdots \\ \sigma^2_{x_{tp}}\end{bmatrix} \qquad \ldots\ldots(22)$$

are the $(m+n+1)\times(1)$ and $(p)\times(1)$ vectors of variances in the frequency and time domain respectively.

and

$$[\sigma_A{}^2] \triangleq \begin{bmatrix}\sigma_{a_1}{}^2 \\ \sigma_{a_2}{}^2 \\ \vdots \\ \sigma_{bm}{}^2\end{bmatrix}, \qquad \ldots\ldots(23)$$

$$\left[\left(\frac{\partial x_f}{\partial A}\right)^2\right] \triangleq \begin{bmatrix}\left(\frac{\partial x_{f1}}{\partial a_1}\right)^2 & \cdots & \left(\frac{\partial x_{f1}}{\partial b_m}\right)^2 \\ \vdots & & \\ \left(\frac{\partial x_{f(m+n+1)}}{\partial a_1}\right)^2, & \cdots & \left(\frac{\partial x_{f(m+n+1)}}{\partial b_m}\right)^2\end{bmatrix} \qquad (24)$$

with a similar definition for $\left[\left(\frac{\partial x_t}{\partial A}\right)^2\right]$ except that it has p rows.

Then it can be shown[21] that

$$[\sigma^2_{x_f}] = \left[\left(\frac{\partial x_f}{\partial A}\right)^2\right][\sigma^2_A] \qquad \ldots\ldots(25)$$

and

$$[\partial^2_{x_t}] = \left[\left(\frac{\partial x_t}{\partial A}\right)^2\right][\sigma^2_A] \qquad \ldots\ldots(26)$$

Eliminating $[\sigma_A{}^2]$ between (25) and (26) then,

$$[\sigma^2_{x_t}] = \left[\left(\frac{\partial x_t}{\partial A}\right)^2\right]\left[\left(\frac{\partial x_f}{\partial A}\right)^2\right]^{-1}[\sigma^2_{x_f}] \qquad \ldots\ldots(27)$$

Equation (27) is the fundamental equation for estimating variances in time domain parameters for known variances in frequency domain parameters.

Conversely if it is required to estimate frequency domain variances from known variances in the time domain parameters then if $[\sigma^2_{xt}]$ and $[\sigma^2_{xf}]$ are now $(r)\times(1)$ and $(m+n+1)\times(1)$ vectors respectively.

$$[\sigma^2_{x_f}] = \left[\left(\frac{\partial x_f}{\partial A}\right)^2\right]\left[\left(\frac{\partial x_t}{\partial A}\right)^2\right]^{-1}[\sigma^2_{xt}] \qquad \ldots\ldots(28)$$

provided the time domain parameters are chosen so that $\left[\left(\frac{\partial x_t}{\partial A}\right)^2\right]^{-1}$ exists and $\left[\left(\frac{\partial x_f}{\partial A}\right)^2\right]$ now has r rows.

16. Appendix 2: An example showing how amplitude ratio variances are predicted from time domain data

The problem is to constrain a family of systems represented by the low-order model of equation (2) to have the following time domain characteristics,

$$\bar{I}_p = 34{\cdot}5 \qquad \bar{S}_p = 1{\cdot}296 \qquad \bar{R}_p = 0{\cdot}0194$$
$$\sigma_{I_p} = 2{\cdot}2 \qquad \sigma_{S_p} = 0{\cdot}040 \qquad \sigma_{R_p} = 0{\cdot}0012$$

by defining gate widths (in terms of variances) on amplitude ratios at spot test frequencies of 18, 50 and 100 radians/second. Figure 8 establishes the coefficients $b = 2{\cdot}9$, $c = 2{\cdot}55$, and Fig. 5 gives $\omega_0 = 46{\cdot}3$.

From equation (26) making full use of normalization frequency ω_0, we have,

$$\left.\begin{aligned}
\frac{1}{\omega_0^2}\sigma^2_{I_p} &= \left(\frac{1}{\omega_0^2}\frac{\partial I_p}{\partial a_1}\right)^2\omega_0^2\sigma^2_{a_1} + \left(\frac{1}{\omega_0^3}\frac{\partial I_p}{\partial a_2}\right)^2\omega_0^4\sigma^2_{a_2} + \\
&\quad + \left(\frac{\partial I_p}{\partial \omega_0}\right)^2\frac{1}{\omega_0^2}\sigma^2_{\omega_0} \\
\sigma^2_{S_p} &= \left(\frac{1}{\omega_0}\frac{\partial S_p}{\partial a_1}\right)^2\omega_0^2\sigma^2_{a_1} + \left(\frac{1}{\omega_0^2}\frac{\partial S_p}{\partial a_2}\right)^2\omega_0^4\sigma^2_{a_2} + \\
&\quad + \left(\omega_0\frac{\partial S_p}{\partial \omega_0}\right)^2\frac{1}{\omega_0^2}\sigma^2_{\omega_0} \\
\omega_0^2\sigma^2_{R_p} &= \left(\frac{\partial R_p}{\partial a_1}\right)^2\omega_0^2\sigma^2_{a_1} + \left(\frac{1}{\omega_0}\frac{\partial R_p}{\partial a_2}\right)^2\omega_0^4\sigma^2_{a_2} \\
&\quad + \left(\omega_0\frac{\partial R_p}{\partial \omega_0}\right)^2\frac{1}{\omega_0^2}\sigma^2_{\omega_0}
\end{aligned}\right\} \ldots(29$$

Similarly, equation (25) may also be written in the form

$$\left.\begin{aligned}
\sigma^2_{x_f} &= \left(\frac{1}{\omega_0}\frac{\partial x_f}{\partial a_1}\right)^2\omega_0^2\sigma^2_{a_1} + \left(\frac{1}{\omega_0}\frac{\partial x_f}{\partial a_2}\right)^2\omega_0^4\sigma^2_{a_2} \\
&\quad + \left(\omega_0\frac{\partial x_f}{\partial \omega_0}\right)^2\frac{1}{\omega_0^2}\sigma^2_{\omega_0}
\end{aligned}\right\} \ldots(30)$$

for $f = 18, 50, 100$ radians/second.

Eliminating $\omega_0^2\sigma^2_{a_1}$, $\omega_0^4\sigma^2_{a_2}$ and $\sigma_0^2\omega_0^2$ between (29) and (30) we obtain equation (8).

The partial derivatives required for evaluating C and D are given in Table 7, the time domain matrix elements being evaluated from Table 3, and the relevant coefficient plane contours. Frequency domain elements are obtained by differentiating the expression for amplitude ratio with respect to the appropriate variable and then substituting spot frequencies.

Table 7
C matrix elements

f	$\frac{1}{\omega_0}\frac{\partial x_f}{\partial a_1}$	$\frac{1}{\omega_0^2}\frac{\partial x_f}{\partial a_2}$	$\omega_0\frac{\partial x_f}{\partial \omega_0}$
18	−0·1448	0·09275	0·1693
50	0·1489	−0·3658	−1·7437
100	0·1359	−0·1376	−0·4356

D matrix elements

x_t	$\partial x_t/\partial a_1$	$\partial x_t/\partial a_2$	$\partial x_t/\partial \omega_0$
I_p	$0{\cdot}195\omega_0^2$	$-0{\cdot}2\omega_0^2$	0·097
S_p	$-0{\cdot}14\omega_0$	$-0{\cdot}01\omega_0$	$(-0{\cdot}415)/\omega_0$
R_p	−0·1974	$0{\cdot}1453\omega_0$	$(-0{\cdot}578)/\omega_0^2$

From Table 7 it is found that

$$CD^{-1} = \begin{bmatrix} 0{\cdot}49873, & 1{\cdot}91789, & -0{\cdot}9176 \\ -9{\cdot}28079, & -28{\cdot}69579, & 24{\cdot}15717 \\ -0{\cdot}10838, & -1{\cdot}04867, & 1{\cdot}11170 \end{bmatrix}$$

so that

$$\sigma^2_{x_{18}} = \frac{-0{\cdot}49873}{(46{\cdot}3)^2}\,\sigma^2_{I_p} + 1{\cdot}91798\sigma^2_{S_p} - 0{\cdot}9176\times \times(46{\cdot}3)^2\sigma^2_{R_p}$$

$$\sigma^2_{x_{50}} = \frac{-9{\cdot}28079}{(46{\cdot}3)^2}\,\sigma^2_{I_p} - 28{\cdot}69579\sigma^2_{S_p} + 24{\cdot}15717\times \times(46{\cdot}3)^2\sigma^2_{R_p}$$

$$\sigma^2_{x_{100}} = \frac{-0{\cdot}10838}{(46{\cdot}3)^2}\,\sigma^2_{I_p} - 1{\cdot}04867\sigma^2_{S_p} + 1{\cdot}11170\times \times(46{\cdot}3)^2\sigma^2_{R_p}$$

After rounding off the coefficients equation (9) is obtained. Putting $\sigma_{I_p} = 2{\cdot}2$, $\sigma_{S_p} = 0{\cdot}040$, $\sigma_{R_p} = 0{\cdot}0012$ the results given in Table 6 are obtained.

Manuscript first received by the Institution on 9th December 1969 and in revised form on 15th June 1971. (Paper No. 1426/IC56).

The Authors

Mrs. J. M. Brown was a Mathematical Scholar of Girton College, Cambridge, and after graduating became head of the mathematical department of a girls' school. Subsequently she did research and development work in the aircraft and guided weapons field with the Hawker Siddeley Group in England and Canada, where she was seconded to the Canadian Armament Research and Development Establishment. In 1959 she became a lecturer in mathematics at the University of Aston and in 1962 took up her present appointment lecturing in mathematics and control engineering at the University of Wales Institute of Science and Technology, where she is also a member of the Dynamic Analysis Group under Professor D. R. Towill.

Professor D. R. Towill (Fellow 1970) was appointed to the Chair of Engineering Production in the Department of Mechanical Engineering and Engineering Production in the University of Wales Institute of Science and Technology at the beginning of 1970. He is a graduate of the Universities of Bristol and Birmingham and before joining UWIST in 1969 was a senior lecturer in automatic control at the Royal Military College of Science. He has held a number of industrial consultancies and was for some years with the then Bristol Aircraft Company as a dynamic analyst.

Professor Towill is the author of numerous papers and books and has received Institution Premiums on three occasions.

Mr. P. A. Payne (Graduate 1965) received his technical education at Willesden Technical College and Borough Polytechnic, London. He has held appointments as an electronic engineer with G. and E. Bradley Ltd., E.M.I. Electronics Ltd. and K and N Electronics Ltd. In 1967 he joined the research staff of the dynamic analysis group in the Department of Mechanical Engineering and Engineering Production of the University of Wales Institute of Science and Technology. Mr. Payne is the author and co-author of a number of papers in the field of dynamic analysis and is currently completing the requirements for a Ph.D. The J. Langham Thompson Premium for 1970 was awarded for a paper by himself, Professor Towill and another colleague.

U.D.C. 621.397.611: 681.323

A Novel Digital Method of Generating a Circle Test Pattern for Television

P. LAPPALAINEN,
Dipl.Eng.*

and

Acting Professor L. OJALA,
Lic. Tech.*

SUMMARY

Numerical computation of the coordinates of the circle points needed is based on simple incremental algorithms. The organization of the special-purpose computer is described. A major advantage over the analogue methods commonly used is the accuracy and the stability of the pattern.

* *Applied Electronics Laboratory, Department of Electrical Engineering, Helsinki University of Technology, Otanemi, Finland.*

1. Introduction

One of the most important patterns for a television test generator is the circle pattern. In studying the geometrical quality and linearity of the television picture all imperfections can be recognized on the basis of the deformations in the test circle.

High accuracy and stability are required of the test circle generation method to achieve reliable results. It is quite difficult to satisfy these requirements by using analogue generation methods or flying-spot scanner techniques;[1] an interpretation is often needed to determine which deformations of the circle arise from deflexion non-linearities and which from the distortion of the initial pattern.

In this paper a digital-oriented alternative is described. The required accuracy and stability of the operation can be achieved easily if the pattern is generated by using computing techniques. The algorithms needed in the computation and the corresponding computer organization are given. The design is based on the use of m.s.i. TTL-circuitry. The latest progress in the manufacturing technology of these circuits makes the construction also economically very attractive.

2. Outlines of Circle Generation Procedure

The method of image scanning on a television screen is presented in simplified form in Fig. 1. The scanning is executed as one field with no interlace. Circle points B_v and D_v on line L_v $(v=1, 2, \ldots, 2R+1)$ are the in and out points of the white insert respectively. The locations of the in and out points are measured from the beginning of each line as multiples of a unit length. This is chosen to be the constant distance between adjacent lines, resulting in equal resolution in horizontal and vertical directions in the circle writing. The circle radius in Fig. 1 is then R units.

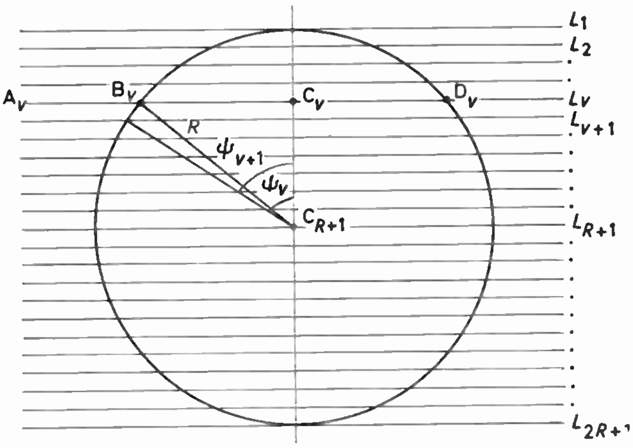


Fig. 1. Image scanning procedure on television screen.

The position of B_v and D_v, measured from the beginning of the line L_v, can be determined according to (1a) and (1b):

$$A_vB_v = A_vC_v - B_vC_v \tag{1a}$$

$$A_vD_v = A_vC_v + B_vC_v \tag{1b}$$

The lengths of the line segments B_vC_v can be computed using simple incremental algorithms. A recursive algorithm based only on computations with integers can be derived to connect the squared values of the lengths of successive line segments to each other. The necessary square root extraction can be performed easily by generating recursively a square sequence $k^2(k = 1, 2, \ldots, R)$ and choosing by comparison the value of k_v^2 being closest to the radicand; k_v is then the best integer-valued approximation for the length of the line segment B_vC_v.

3. Derivation of Basic Recursive Algorithms

By studying the rectangular triangles $B_vC_vC_{R+1}$ and $B_{v+1}C_{v+1}C_{R+1}$, (2a) and (2b) can be obtained.

$$(R \sin \psi_v)^2+(R-v+1)^2 = R^2 \qquad \ldots\ldots(2a)$$

$$(R \sin \psi_{v+1})^2+(R-v)^2 = R^2 \qquad \ldots\ldots(2b)$$

where

$$v = 1, 2, \ldots, 2R.$$

Subtracting (2a) from (2b) and defining

$$y_v = (R \sin \psi_v)^2, \qquad \ldots\ldots(3)$$

yields recursion formula (4).

$$y_{v+1} = y_v+(2R+1-2v). \qquad \ldots\ldots(4)$$

where

$$y_1 = 0.$$

The recursion formula (4) connects the squared values of the lengths of successive line segments to each other in the simplified system given in Fig. 1. All the principal television systems, however, utilize 2-to-1 interlaced scanning, and therefore the situation is somewhat more complicated.

By a similar derivation as before, corresponding results (5) and (6) can be achieved for odd and even fields, respectively:

$$y_{2v+1} = y_{2v-1}+(4R+4-8v) \qquad \ldots\ldots(5)$$

where

$$v = 1, 2, \ldots, R.$$

and

$$y_1 = 0.$$

$$y_{2v+2} = y_{2v}+(4R-8v) \qquad \ldots\ldots(6)$$

where

$$v = 1, 2, \ldots, R-1$$

and

$$y_2 = 2R-1.$$

These results, however, can further be simplified by generating the increments in (5) and (6) recursively. This can be seen from Table 1, where the structures of increment sequences are given for odd and even fields, respectively.

The successive increments differ from each other by the integer eight. Utilizing this fact, and changing number notation into signed two's complement representation, the following final algorithms can be written for odd and even fields, respectively:

$$y_{2v+1} = y_{2v-1}+\Delta y_{2v-1} \qquad \ldots\ldots(7a)$$

where

$$v = 1, 2, \ldots, R,$$

$$y_1 = 0$$

and

$$\Delta y_{2v+3} = \Delta y_{2v+1}+\overline{(-8)} \qquad \ldots\ldots(7b)$$

where

$$v = 1, 2, \ldots, R-1,$$

$$\Delta y_3 = 4R-4.$$

$$y_{2v+2} = y_{2v}+\Delta y_{2v+2} \qquad \ldots\ldots(8a)$$

where

$$v = 1, 2, \ldots, R-1,$$

$$y_2 = 2R-1$$

and

$$\Delta y_{2v+4} = \Delta y_{2v+2}+\overline{(-8)} \qquad \ldots\ldots(8b)$$

where

$$v = 1, 2, \ldots, R-2,$$

$$\Delta y_4 = 4R-8.$$

The bar in these formulas represents the complementing operation.

Table 1. Increment sequences

Field			
Odd		Even	
$2v+1$	$\Delta y_{2v+1} = y_{2v+1}-y_{2v-1}$	$2v+2$	$\Delta y_{2+2} = y_{2v+2}-y_{2v}$
3	$4R-4$	4	$4R-8$
5	$4R-12$	6	$4R-16$
.	.	.	.
.	.	.	.
.	.	.	.
R	$+8$	$R+1$	$+4$
$R+2$	0	$R+3$	-4
$R+4$	-8		
.	.	.	.
.	.	.	.
.	.	.	.
$2R-1$	$-(4R-12)$	$2R-2$	$-(4R-16)$
$2R+1$	$-(4R-4)$	$2R$	$-(4R-8)$

4. Square Root Extraction Algorithm

In this section an elementary and economic method for square root extraction is studied and applied to the circle generation procedure.

When an integer-valued approximation k_v of the square root of an integer-valued radicand y_v must be found, a very natural method is:

(1) generate a square sequence k^2 $(k = 1, 2, \ldots)$,

(2) compare every square value one after the other with the radicand y_v,

(3) detect the first value of k_v^2, which exceeds or equals to y_v, and

(4) choose the corresponding k_v as the result.

The upper approximate value of the square root has been determined.

The square of an integer k is obtained according to Braun[2] from eqn. (9):

$$S_k = k^2 = \sum_{i=1}^{k} a_i = \sum_{i=1}^{k} (2i-1)$$
$$= 1+3+5+ \ldots +(2k-3)+(2k-1) \qquad \ldots\ldots(9)$$

As can be seen, each addend in (9) has the value of its predecessor incremented by the integer two. Hence, (9) can be constituted by a pair of recursive incremental algorithms (10) and (11) which are very suitable for square generation, because the mechanization of the corresponding computations is quite simple:

$$a_{i+1} = a_i+2 \qquad \ldots\ldots(10)$$

where $a_1 = 1$.

$$S_{i+1} = S_i+a_{i+1} \qquad \ldots\ldots(11)$$

where $S_1 = 1$.

The use of the upper approximate value of the square root (truncating algorithm) produces a systematic round-off error. This error, however, can be reduced considerably if properly rounded values are used. This can be achieved by slightly modifying the preceding procedure as is shown in Appendix 1.

In the circle generation procedure the situation is more complicated. The length of the line segment B_vC_v achieves its maximum value of R units when the angle ψ (Fig. 1) equals 90°. Thereafter the lengths begin to decrease and square root extraction of decreasing number sequence will be executed. The decreasing square root sequence needed in the extraction corresponding to (9) can be generated according to (12):

$$S_{R-k} = (R-k)^2 = R^2 - \sum_{i=1}^{k} b_i = R^2 - \sum_{i=1}^{k} [2R-(2i-1)]$$
$$= R^2-(2R-1)-(2R-3)- \ldots -[2R-(2k-1)] \qquad \ldots\ldots(12)$$

An equivalent pair of recursive incremental algorithms is given in (13) and (14):

$$b_{i+1} = b_i-2 \qquad \ldots\ldots(13)$$

where $b_1 = 2R-1$.

$$S_{R-i} = S_{R-i+1}-b_i \qquad \ldots\ldots(14)$$

where $S_R = R^2$.

For computing purposes, it is necessary to transform the preceding equations into signed two's complement form. Thus

$$(\overline{-b_{i+1}}) = (\overline{-b_i})+2 \qquad \ldots\ldots(15)$$
$$S_{R-i} = S_{R-i+1}+(\overline{-b_i}) \qquad \ldots\ldots(16)$$

When S_{R-i} falls below the radicand y_v, a lower approximate value of the square root is obtained.

5. Organization and Operation of the Computing Unit

A computing test circle generator has been designed to satisfy the requirements of C.C.I.R. television system B.[3] In this system there are 625 lines, forty of which belong to the field blanking period. Thus the number of lines which can be seen is 585. The diameter of the circle to be generated has been chosen to be 75% of the height of the picture; R, the radius of the circle, is then 215 units.

In Fig. 2 the main computer organization is given. There are additional devices belonging to the generator, e.g. field detector, position counters, interpreting logic, etc., but because of their minor importance they are not presented here.

Operational unit I generates according to (7a) and (7b) or (8a) and (8b) the squares of the lengths of the line segments B_vC_v. Operational unit II generates according to (10) and (11) or (15) and (16) the required square sequence of integers. Square root extraction is completed with a binary comparator, which makes the decision of interruption, when the condition (C) $\leq$ (F) for $\psi \leq 90°$ or (C) $\geq$ (F) for $\psi > 90°$ is satisfied.

5.1. *Operation Details*

At the specific line m of the field, where the circle writing on television screen begins, Q_6 at the J-input of JK-FF1 becomes true. Q_6 is a logical variable which determines the vertical position of the circle. Thereafter Q_1 and Q_2 become true, and clock pulses can reach operational unit I which computes serially the square of the line segment $B_{m+1}C_{m+1}$ into the register C. The square of the line segment B_mC_m is given as the initial value of register C for each field. When the contents of register B, which is constant $(\overline{-8})$, has recirculated to its original position, the operation of this unit must be halted. This condition is detected by the decoding gate P_3 and the operation interrupted by JK-FF2.

Then the operation of unit II is initiated through the gate P_4 and JK-FF3. Clock pulses can reach the unit which serially generates squares of integers in sequence into register F. When the contents of register D, which is constant 2, has recirculated to its original position, the operation must be halted. This is done by JK-FF3.

Then the binary comparator compares the contents of registers C and F. If (C) > (F) JK-FF3 is set again, and a new cycle is executed. This action continues until (C) $\leq$ (F). The up/down counter counts the total number of execution cycles of unit II, which is a measure of the length of $B_{m+1}C_{m+1}$. Q_5 in the comparison block is a logical variable which is false if $\psi \leq 90°$, and true if $\psi > 90°$.

The next line pulse sets JK-FF2 and the action as described is repeated. The whole procedure is valid for $0 < \psi \leq 90°$. Thereafter the lengths of the line segments begin to decrease and the mode of operation must be changed. Then unit I generates squares of the lengths of the line segments as earlier, but until II generates decreasing square sequence of integers, and the comparator detects the condition (C) $\geq$ (F). Otherwise the operation is the same as described.

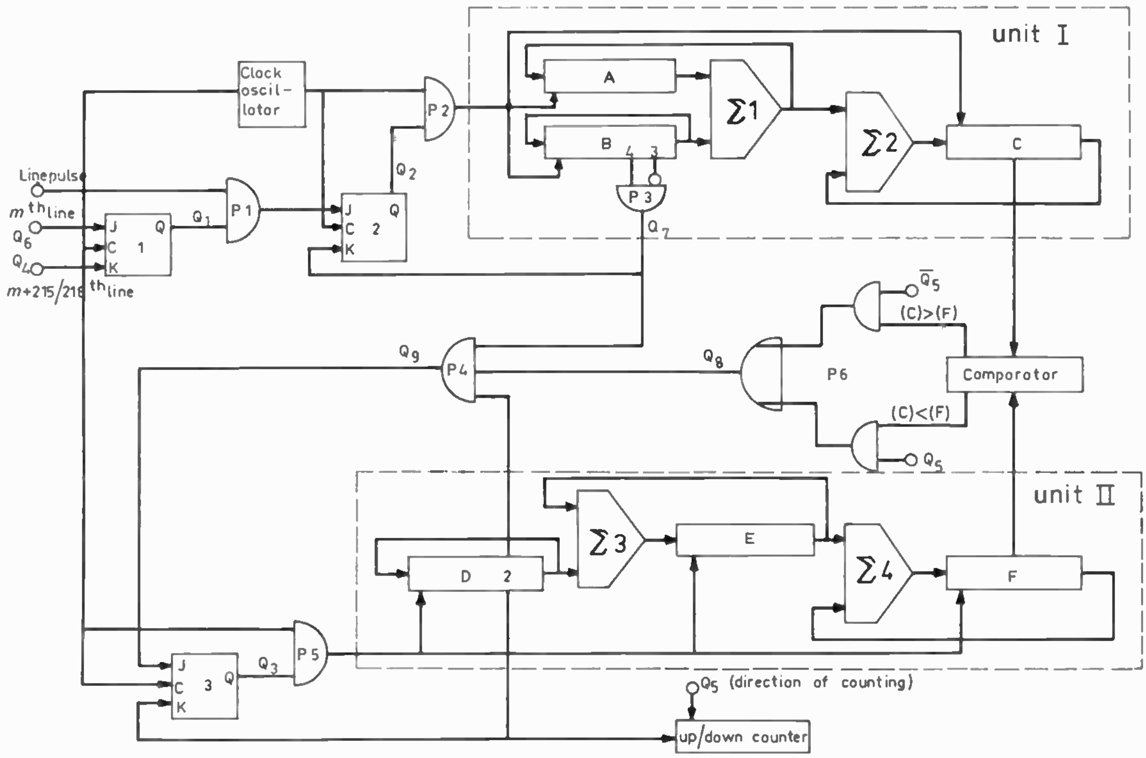


Fig. 2. Main computer organization.

At line $m+215$ or $m+216$, depending on the field, Q_4 at the K-input of JK-FF1 becomes true, and so the operation is interrupted for the rest of the field.

In Appendix 2 a symbolic representation of line information computation has been given based on the notational formalism of Chu.[4]

5.2. *Initial Values*

The initial values of shift registers A, B, and C in Fig. 2 must be chosen according to equations (7) and (8) as given in Table 2 for odd and even fields respectively.

Table 2. Initial values for computation

Register	Field	
	Odd	Even
A	$4R+4=864$	$4R=860$
B	$\overline{(-8)}$	$\overline{(-8)}$
C	0	$2R-1=429$

Initial values of shift registers D, E, and F in Fig. 2 can be obtained from (10) and (11). They are, respectively, 2, 1, and 0.

If these values are used, the clock frequency must be quite high (13 MHz) to find time to compute the coordinates of next white insert during one line period. In order to reduce the clock frequency, it is reasonable to compute beforehand the coordinates needed for the first lines of both fields and to appropriate initial values. By this means, the clock frequency can be nearly halved. The initial values needed are given in Table 3 for odd and even fields, respectively.

Table 3. Initial values for square sequence generation

Register	Field	
	Odd	Even
D	2	2
E	59	43
F	900	441

When ψ exceeds 90°, a decreasing square sequence is generated on the basis of (15) and (16). Then the sign bit in shift register E must be changed and its contents represented in two's complement form: $\overline{(-429)}$.

5.3. *Simulation*

The circle generation procedure based on the truncating square root extraction algorithm (9) as formerly described and on initial values given in Tables 2 and 3 has been simulated with a general-purpose computer. In Fig. 3 the left half of the resulting test circle is shown. The circumference is smooth enough for practical testing purposes. In Fig. 4 the region of circumference around 90° is shown magnified; the asymmetrical round-off error mechanism can clearly be seen.

5.4. *Critical Considerations*

Another digital-oriented construction for circle pattern generation was presented by Philips in 1968 in Montreux. The generator is based on the use of a ferrite core read-only memory where the needed information for circle writing is stored. The performance of the device is comparable with the computing circle generator presented in this paper.

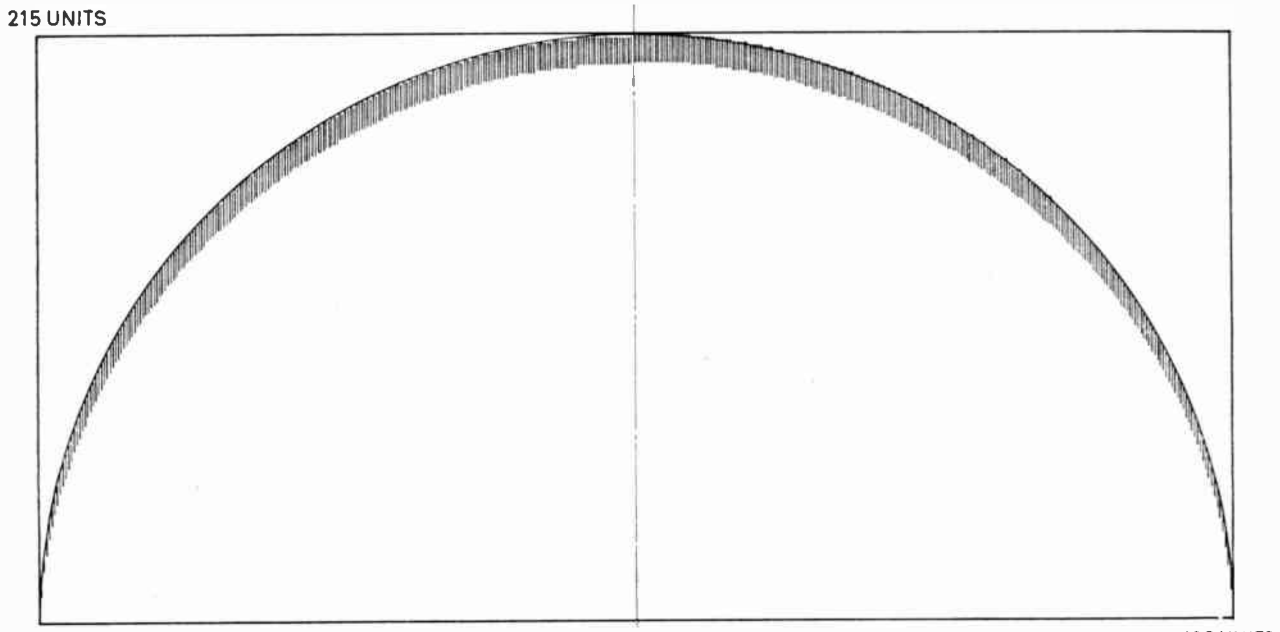


Fig. 3. Left half of the generated test circle.

The computing circle generator, however, has the advantage of standard TTL-circuitry instead of rather complex and expensive read/write electronics of the ferrite core r.o.m. generator. Preliminary investigations concerning the use of switching matrices as the source of line information for circle test pattern generation have also been started by one of the authors.[5] However, so far as the authors of this paper know there are no other digital circle generators available.

6. Conclusions

In this paper a new method for the television test circle generation has been proposed. The positions of the points on the circumference are determined by numerical computations based on simple incremental algorithms. The corresponding computer configuration has also been designed.

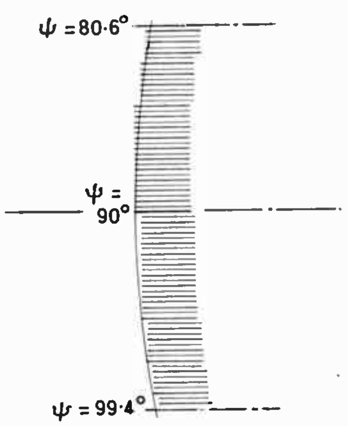


Fig. 4. Structure of the test circle circumference around $\psi = 90°$ (truncating algorithm).

The precision and stability of the method are excellent: the algorithms are exact integer algorithms without any truncation or rounding error accumulation. This property is very favourable especially for extensive computations. For problems of medium size like the one described here, the use of read-only memories for the storage of line information seems to be quite competitive.

7. Acknowledgments

The authors are grateful to Professor P. Jääskeläinen for his interest in this work and encouraging attitude. Their gratitude is also directed to Dipl.Eng. V. Huttunen and Dipl.Eng. T. Leinonen for helpful discussions, and the Foundation of Technology in Finland for its economical support.

8. References

1. Hazeltine Staff, 'Principles of Color Television' (John Wiley, New York, 1956).
2. Braun, E., 'Digital Computer Design' (Academic Press, New York, 1963).
3. C.C.I.R., Documents of the XIth Plenary Assembly, Oslo 1966, vol. V (International Telecommunication Union, Geneva, 1967).
4. Chu, Y., 'Digital Computer Design Fundamentals' (McGraw-Hill, New York, 1962).
5. Lappalainen, P., 'Digital Circle Pattern Generator for TV', VIIth Symposium on Radio Science, 1970. U.R.S.I., Finnish National Committee, Paper 13.4, 3 pages (in Finnish).
6. Cliff, R., 'Simplifying square-root computation', *Instruments and Control Systems*, **41**, pp. 114–5, April 1968 (NASA Tech. Brief 65-10343).

9. Appendix 1: Improved square root extraction algorithm

In the preceding square root extraction procedure a truncated approximate value of the square root was obtained. The resulting systematic round-off error can be reduced considerably if properly rounded values are used.

When an integer-valued approximation k_v of the square root of an integer-valued radicand y_v is to be found, a method producing rounded values is:[4]

(1) generate a sequence k^2+k $(k = 1, 2, \ldots)$,
(2) compare every value of this sequence one after the other with the radicand y_v,
(3) detect the first value of k_v^2, which exceeds or equals y_v, and
(4) choose the corresponding k_v as the results.

The integer sequence (k^2+k) can be generated incrementally from:

$$S'_k = \sum_{i=1}^{k} a'_i = \sum_{i=1}^{k} 2i = 2+4+6+\ldots+2k = k^2+k \qquad \ldots\ldots(17)$$

An equivalent pair of recursive incremental algorithms is given in (18) and (19):

$$a'_{i+1} = a'_i+2 \qquad \ldots\ldots(18)$$

where $a'_1 = 2$

and

$$S'_{k+1} = S'_k+a'_{k+1} \qquad \ldots\ldots(19)$$

where $S'_1 = 2$.

When this improved square root extraction algorithm is applied to the circle generation procedure, (17) to (19) are directly applicable, when the angle $\psi \leq 90°$. When S'_k exceeds or equals the radicand, the rounded value of the square root is obtained.

In the generation of the decreasing number sequence ($\psi > 90°$) needed in the procedure, (12) and (16) must be replaced with corresponding modified algorithms (20) to (24), respectively. Thus:

$$S'_{R-k} = (R-k)^2+R-k = R^2+R-\sum_{i=1}^{k} b_i$$

$$= R^2+R-\sum_{i=1}^{k}(2R-(2i-2)) \qquad \ldots\ldots(20)$$

$$= R^2+R-2R-(2R-2)-(2R-4)-\ldots-$$
$$-(2R-(2k-2))$$

$$b_{i+1} = b_i-2 \qquad \ldots\ldots(21)$$

where $b_1 = 2R$

$$S'_{R-i} = S'_{R-i+1}-b_i \qquad \ldots\ldots(22)$$

where $S'_R = R^2+R$

$$(\overline{-b_{i+1}}) = (\overline{-b_i})+2 \qquad \ldots\ldots(23)$$

$$S'_{R-i} = S'_{R-i+1}+(\overline{-b_i}) \qquad \ldots\ldots(24)$$

When S'_{R-i} falls below the radicand, the rounded value of the square root must be taken as $R-i+1$.

The circle generation procedure based on the rounding square root algorithm and on corresponding initial values has been simulated with a general-purpose computer. For comparison the structure of the test circle circumference around the same angle region as in Fig. 4 is given in Fig. 5.

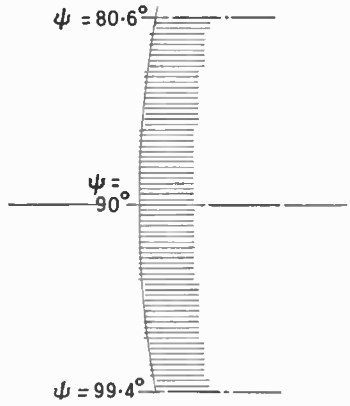


Fig. 5. Structure of the test circle circumference around $\psi = 90°$ (rounding algorithm).

9.1. *Round-off Procedures*

The basic square root extraction algorithms described have different round-off properties. The mappings between integer-valued radicands and integer-valued square root approximations are given in Fig. 6 for both algorithms. It can be seen that the rounding algorithm is also optimal by studying the identities (25) and (26) and Fig. 6.

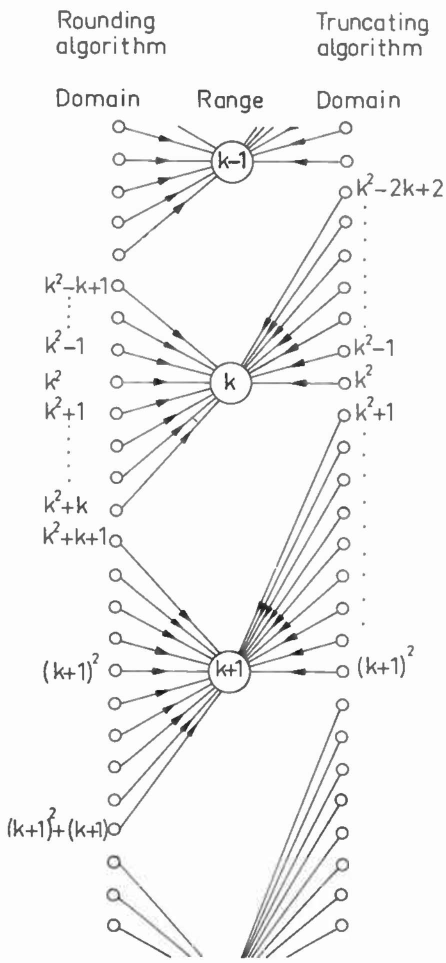


Fig. 6. Mapping of the square root function.

$$(k+1/2)^2 = [(k+1)-1/2]^2 = (k^2+k)+1/4 \quad \ldots\ldots(25)$$

$$(k-1/2)^2 = [(k-1)+1/2]^2 = (k^2-k)+1/4 \quad \ldots\ldots(26)$$

An obvious way to compare these round-off procedures is to compute the error statistics when the resulting value of the square root extraction is k. The results concerning mean and variance are given in (27) to (30).

Rounding algorithm

Mean

$$\eta_k = \frac{1}{2k}\sum_{v=-k+1}^{k}(\sqrt{k^2+v}-k) \xrightarrow[k \gg 1]{} 1/4k \xrightarrow[k\to\infty]{} 0 \quad \ldots\ldots(27)$$

Variance

$$\sigma_k^2 = \frac{1}{2k}\sum_{v=-k+1}^{k}(\sqrt{k^2+v}-k-\eta_k)^2 \xrightarrow[k\to\infty]{} 1/12 \quad \ldots\ldots(28)$$

Truncating algorithm

Mean

$$\eta_k = \frac{1}{2k-1}\sum_{v=-(2k-2)}^{0}\sqrt{k^2+v}-k) \xrightarrow[k\gg\infty]{} \frac{1}{2}+\frac{1}{2k} \xrightarrow[k\to 1]{} -\frac{1}{2} \quad (29)$$

Variance

$$\sigma_k^2 = \frac{1}{2k-1}\sum_{v=-(2k-2)}^{0}(\sqrt{k^2+v}-k-\eta_k)^2 \xrightarrow[k\to\infty]{} 1/12 \quad (30)$$

The limits are equivalent to the results concerning round-off procedures (rounding and truncation) in ordinary arithmetics.[3]

A prominent feature of these incremental algorithms is, however, that there is no error accumulation. This means that these computational methods based on integer algorithms are quite comparable in numerical performance with r.o.m.-tables, where properly rounded values of functions are tabulated. In applications where very extensive tables are needed these computational methods are superior in economy.

10. Appendix 2: Symbolic representation of line information computation

```
    Q6 : 1 => Q1
1)  Q1 : 1 => Q2
    Q2 : (A) + (B) => A,    (A) + (B) + (C) => C
    Q7 : 0 => Q2
    Q9 : 1 => Q3
2)  Q3 : (D) + (E) => E,    (E) + (F) => F
    D2 : 0 => Q3
    Q8 : 1 => Q3    GO TO 2
    SQUARE ROOT EXTRACTION COMPLETED
    Q̄5 : GO TO 1
3)  Q1 : 1 => Q2
    Q2 : (A) + (B) => A     (A) + (B) + (C) => C
    Q7 : 0 => Q2
    Q9 : 1 => Q3
4)  Q3 : (D) - (E) => E,    (F) - (E) => F
    D2 : 0 => Q3
    Q8 : 1 => Q3,    GO TO 4
    SQUARE ROOT EXTRACTION COMPLETED
    Q̄4 : GO TO 3
    Q4 : 0 => Q1
    END OF EXECUTION
```

Manuscript first received by the Institution on 11th August 1971 and in final form on 10th November 1971. (Paper No. 1427/CC114)

U.D.C. 621.372

Generalized Karnaugh Maps

F. C. EVANS,
B.Sc., M.Inst.P., C.Eng., M.I.E.R.E.*

SUMMARY

The process of mapping Boolean functions is generalized from the Karnaugh map, and it is shown that there are eight useful maps corresponding to any function, from which eight two-layer networks may be derived for implementation of the function. Applications to e.c.l. and m.o.s. logic are mentioned.

**School of Physical Sciences, Department of Physics, University of St. Andrews, North Haugh, Fife. Scotland.*

The well-known Karnaugh map[1] is a method of representing a Boolean function in which the literals corresponding to each cell containing a 1 are combined by means of the connective AND to form the term, and the terms are combined by means of the connective OR to form the function. In the case of the less well-known inverse Karnaugh map,[2] the roles of these connectives are interchanged. In this paper the process of mapping is generalized on the assumption that the literals and the terms may each be combined by means of any of the four connectives AND, OR, NAND or NOR, giving 16 different kinds of map in principle. Each possibility is then examined by means of three postulated tests, to determine whether it is a useful map. The 16 possible kinds of map are listed in Table 1, column 1; the first connective stated is that by which the literals are combined, and the second is that by which the terms are combined.

Table 1

Name of map	Test 1	Type of map or equivalent single gate	Test 2	Test 3
AND/OR	√	K	√	√
NAND/NAND	√	K	√	√
OR/AND	√	I	√	√
NOR/NOR	√	I	√	√
NAND/AND	√	$\bar{K}$	√	√
AND/NOR	√	$\bar{K}$	√	√
NOR/OR	√	$\bar{I}$	√	√
OR/NAND	√	$\bar{I}$	√	√
AND/AND	×	AND		
OR/OR	×	OR		
AND/NAND	×	NAND		
OR/NOR	×	NOR	Not applied	
NAND/OR	×	NAND		
NAND/NOR	×	AND		
NOR/NAND	×	OR		
NOR/AND	×	NOR		

The three tests postulated in order to decide the usefulness of each kind of map are as follows:

1. It must be possible to plot any logical function.
2. The complement of the function should correspond to the negated map, i.e. the map formed by interchanging 1s and 0s.
3. Grouping of individual cells into 2s, 4s, 8s, etc., must be meaningful for minimization purposes.

It is known that the ordinary Karnaugh map (the AND/OR) and the inverse Karnaugh map (the OR/AND) satisfy these conditions. The other kinds of map may be tested by starting with a general function in the form corresponding to each kind of map, and attempting to manipulate it algebraically into an AND/OR or an OR/AND function. All the proofs are similar, so for the sake of brevity only a few examples are given but all the results are tabulated in Table 1. De Morgan's theorem is used for the manipulation in every case.

Results of test 1

(a) NAND/NAND map

The general form of function is

$$F=\overline{\overline{(A_1B_1C_1\ldots)}\;\overline{(A_2B_2C_2\ldots)}\ldots}$$

which reduces to

$$F=A_1B_1C_1\ldots+A_2B_2C_2\ldots+\ldots$$

The function is therefore identical with the AND/OR function using the same literals, and the map is identical with the AND/OR, or Karnaugh map. (This is in accordance with the well-known fact that an AND/OR network may be replaced by a NAND network with which it is topologically identical.) Any logical function can therefore be plotted on a NAND/NAND map.

(b) NOR/NOR map

The general form of function is

$$F=\overline{\overline{(A_1+B_1+C_1+\ldots)}+\overline{(A_2+B_2+C_2+\ldots)}+\ldots}$$

which may be written

$$F=(A_1+B_1+C_1\ldots)(A_2+B_2+C_2+\ldots)\ldots$$

The map is therefore identical with the OR/AND or inverse Karnaugh map, and again any logical function can be plotted.

(c) AND/NOR map

The general form of function is

$$F=\overline{A_1B_1C_1\ldots+A_2B_2C_2\ldots+\ldots}$$
$$=(\bar{A}_1+\bar{B}_1+\bar{C}_1+\ldots)(\bar{A}_2+\bar{B}_2+\bar{C}_2+\ldots)\ldots$$

i.e. an OR/AND function in terms of the negated literals. The map is therefore a negated Karnaugh map, i.e. the result of plotting a Karnaugh map for the same set of literals and then interchanging 1s and 0s.

(d) NAND/OR map

The general form of function is

$$F=\overline{A_1B_1C_1\ldots}+\overline{A_2B_2C_2\ldots}+\ldots$$
$$=\overline{A_1B_1C_1\ldots A_2B_2C_2\ldots}$$

i.e. a single NAND term. It is not therefore possible to plot any logical function on a NAND/OR map, but only those which correspond to a single NAND term. It also follows that a two-layer NAND/OR network of gates is equivalent to a single NAND gate, and such a network may be used to build a NAND gate with any required number of inputs.

When test 1 is applied to all the possible kinds of map, eight of them pass the test and eight fail. Those which pass the test comprise two each of the following types, Karnaugh, inverse Karnaugh, negated Karnaugh and negated inverse Karnaugh. These are indicated by K, I, $\bar{K}$ and $\bar{I}$ respectively in Table 1, column 3. Those maps which fail test 1 are all found to be equivalent to a single term in the literals represented, and the nature of this term for each case is indicated in Table 1, column 3.

Results of tests 2 and 3

Those maps which pass test 1 also pass tests 2 and 3 since: (i) these are known properties for both 1s and 0s in maps type K and type I, and (ii) maps types $\bar{K}$ and $\bar{I}$ are simply formed by interchanging 1s and 0s on maps types K and I respectively. Alternatively, general proofs on the same lines as those used for test 1 are very easily formulated. In a restricted sense, all those maps which fail test 1 might also be said to pass tests 2 and 3, but since these are not useful types of map for minimization purposes tests 2 and 3 have not formally been applied.

Applications

The ordinary Karnaugh map is a very useful design aid for use in connexion with the older types of logic such as DTL and TTL, since these are mainly NAND systems. The minimal two-layer AND/OR network found from this type of map can be converted to a NAND network by direct substitution, and this will usually yield a reasonably economical solution. In the case of the more recent ECL types of logic, the gates commonly provide two outputs, namely OR and NOR functions of the inputs. In this case, the designer can see from Table 1 that, for example, he can if he wishes think in terms of OR/AND logic, carry out his minimization process using a map type I, and then implement the resulting minimal function using a NOR/NOR network. Alternatively a NOR/OR network could be designed by using a map type $\bar{I}$. It is also seen that, for example, an OR/NOR network does not implement general solutions, but may be used to build an expanded NOR gate. The table also shows that systems using the commonly available AND-OR-INVERT, or AND/NOR gates may be designed directly from a map type $\bar{K}$. This type of network is currently of interest because of its availability in m.o.s. logic circuits.

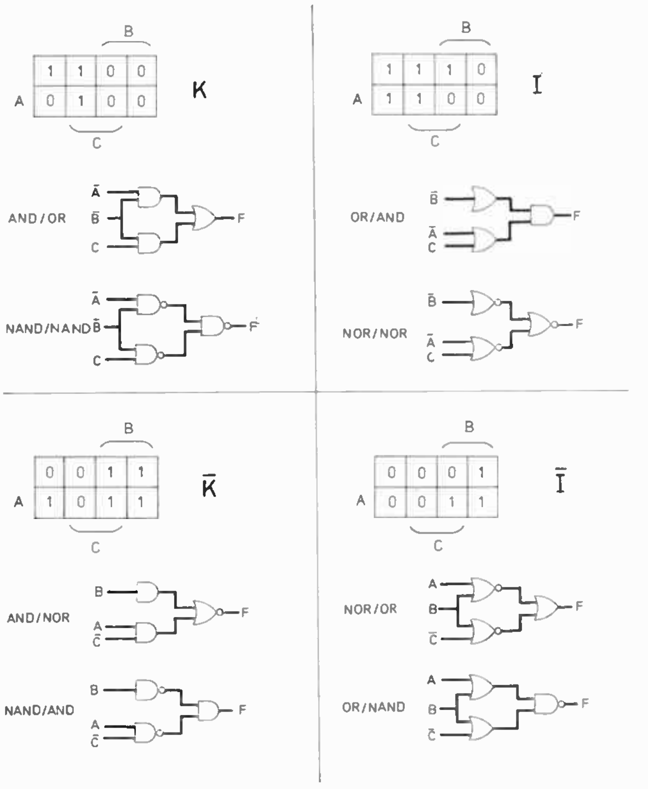


Fig. 1. The four types of Karnaugh map and the two-layer networks derived from them.

Figure 1 shows as an example the four types of map, and the eight two-layer networks derived from them, for the function $F = A\bar{B}C + \bar{A}\bar{B}C + \bar{A}\bar{B}\bar{C}$. Obvious simplifications are possible in those networks which contain single-input non-inverting gates, but the networks have been left in the form which best illustrates the general result.

References

1. Karnaugh, M., 'The map method for synthesis of combinational logic circuits', *Commun. and Electronics*, No. 9, pp. 593–9, November 1953.
2. Evans, F. C., 'Use of inverse Karnaugh maps in realization of logic functions', *Electronics Letters*, 5, No. 21, pp. 537–8, 16th October 1969; Correction, *ibid.*, No. 25, p. 670, 11th December 1969.

Manuscript received by the Institution on 30th June 1971. (Short Contribution 150/CC115.)

LETTERS

Early Submarine Cable Construction

I read with great interest the paper by R. M. Barker, 'A royal intervention in early submarine telegraphy',[1] with its account of Wheatstone's attempts to make a satisfactory submarine cable.

A practical submarine cable only became possible in 1848 when gutta percha was available in quantity and Charles Hancock had solved the problem of extruding a continuous sheath over a wire.[2]

Mr. Barker suggests that the cable Wheatstone tried out in Swansea Bay in the summer of 1844 was insulated with rubber. I think this is unlikely. Even if rubber were available in sufficient quantity at the time, which is doubtful, there would remain the problem of extruding a sheath. It seems far more likely that Wheatstone was testing in the sea the rope and tar insulation which until then had only been tried in the River Thames from King's College where he was Professor of Experimental Philosophy.

Unless any new evidence turns up we cannot be sure what materials Wheatstone used in each of his underwater cable experiments, but I suggest the reconstruction of events which follows. I imagine that all were in practice single-core cables even though the published plans show multi-core.

1. In 1840 Wheatstone announced his plan for a cable whose cores were insulated with rope 'well saturated with boiled tar'.
2. Wheatstone would have experimented with a rope-and-tar insulated cable in the Thames, and probably shown this to Prince Albert when he visited the College in 1843.
3. Wheatstone would have demonstrated this to other eminent people interested in science, including J. D. Llewellyn with whom he conducted the trials at Swansea in 1844. We must ask why Wheatstone travelled so far. Perhaps Llewellyn arranged for a cable to be made—there was a copper works nearby. If Wheatstone had by then made a cable which was satisfactory in the river, he would have wished to try it out at sea to ascertain the effects of salt water and the action of the waves.
4. The sea trials proved a failure, for, as Mr. Barker says, two years later Wheatstone was trying out a lead covered cable at Portsmouth. This cable was insulated with wool or cotton impregnated with marine glue.

After this Wheatstone was not personally concerned with submarine telegraphy for several years, and when he took up the subject again later gutta percha was in general use. His submarine cable work was a series of failures, but it is nevertheless an interesting story of a great scientist's attempt to solve the problem of underwater insulation; a problem that was eventually solved only after an enormous expenditure of scientific effort and hard cash.

Brian Bowers,
B.Sc.(Eng.), C.Eng., M.I.E.E.

Department of Electrical Engineering,
Science Museum, London SW7 2DD.
21st December 1971

Mr. Barker writes in reply:

I doubt whether Wheatstone was still interested in rope and tar insulation in 1844. Cooke described the first Enderby hemp-covered cable as having been intended for submarine work and to have shown its unsuitability by breaking down under a shower of rain when it was in use between Euston and Camden Town in 1837.[3] In view of this failure, it is surprising that Wheatstone should propose using a tarred-rope cable for his cross-channel plans of 1840.

There is some evidence that Wheatstone may have had a rubber-covered cable by the time he carried out the tests in Swansea Bay in 1844. W. T. Henley, who worked at one time for Wheatstone, said, in his evidence to the 1861 Committee, that he made some india rubber-covered wire, 'about 1843, I think'. Henley said that strips of india-rubber were run round the wire, which was then passed through naphtha and compressed, but that it was difficult to make the rubber adhere.

Leopold Martin, in his reminiscences,[4] describes seeing the wire across the Thames from King's College to the shot tower and, at about the same time, accompanying Wheatstone on a visit to Enderby's India Rubber Cable Factory to see the new process of vulcanizing, which Wheatstone proposed using for his submarine telegraph insulation.

Although not conclusive, this evidence leads me to believe that, by 1844, Wheatstone had rejected the tar-and-rope insulation and that rubber was one of the materials he tested. I agree with Mr. Bowers that Wheatstone's submarine cable work was a series of failures, but he did succeed in passing signals from a boat to the lighthouse in Swansea Bay, in 1844. As this was probably the first time that signals had been sent through a cable under the sea, there is some justification for describing the event as the starting point of submarine telegraphy.

[1] *The Radio and Electronic Engineer*, **41**, No. 11, pp. 522–4, November 1971.

[2] Garratt, G. R. M., '100 Years of Submarine Cables,' p.7. (Science Museum, London, 1951.) [Now out of print.]

[3] Hubbard, G., 'Cooke and Wheatstone and the Invention of the Electric Telegraph', p. 48, p. 133. (Routledge and Kegan Paul, London, 1951.)

[4] Martin, L., *Newcastle Daily Chronicle*, March 30th, 1889.

U.D.C. 621.86

A Measurement Transducer and Program Controller for Automatic Mechanical Handling

E. HOWBROOK,
M.Sc., C.Eng., M.I.E.E.*

Based on a paper presented at the Conference on Electronic Control of Mechanical Handling held in Nottingham from 6th to 8th July 1971.

SUMMARY

Manned mechanical handling is gradually being augmented by automatic systems, and it is appropriate to review manufacturing processes in this connexion. The principles and use of linear and rotary electromagnetically-coupled measurement transducers giving the required information in the form of a change of phase of an electrical signal are discussed. The extent to which transducers can be used as the basis of automatic systems for point-to-point transportation as, for example, in automatic warehousing is described.

* *Department of Electrical Engineering, University of Salford, Salford M5 4WT.*

1. Introduction

Since the early 1960s the availability of low cost, reliable computers has made a considerable impact on industrial and commercial activities. It is now commonplace to find computers in banking, accounting, costing and stock-control etc., and, after initial installation difficulties have been overcome, it becomes unthinkable to return to the old methods. The introduction of the computer into manufacturing processes has been slow, due in part to the need to change existing expensive works and equipment gradually as replacement or production dictates. Nevertheless, computer control can be found in the steel industry, rolling mills, warehousing etc.

In order that further progress can continue, it is important to analyse the whole factory process so that, where possible, each small change is made within the framework of the final objective of computer-assisted production.

An overall analysis leads to the conclusion that automatic mechanical handling with computer control is a key feature, and in particular, the system of sensing the position of material or a workpiece, and the movements in response to a program, are vital factors.

2. The Computer and Mechanical Handling

To obtain high productivity in a factory, the transportation of material or parts and storage of these must be undertaken efficiently and reliably. Where production is of a repetitive nature and overall computer control is envisaged, it is essential that the storage transportation, and mechanical handling should be automatically controlled, and the form of the control, the transducers, and program controllers, should be considered together. A computer could, if required, control every individual movement of all mechanical handling. This would require an encoded signal fed back to the computer from every functional part of every machine, leading to excessive wiring and computer work. A more efficient and flexible arrangement would be to restrict the computer to overall planning. For example the computer decision could be 'my stock level of part B is low—issue more material B from store to machine No. 6'.

With the computer limited to the overall planning, the local movement of material can be controlled by methods best suited to each particular case, and where the mechanical handling is to be automatic, a program of movements, determined by a program controller will be initiated by the computer. In this case the position sensors or transducers may have a digital or analogue form depending on the requirements of the control and program controller.

The present paper describes a family of transducers which, it is suggested, are suitable for mechanical handling, and two program controllers which are convenient for mechanical handling and compatible with the transducers and control. An extension of the transducer system is proposed for automatic warehousing or point-to-point transportation.

3. The Transducer

3.1. *Principle*

The variable-phase transducer is an attempt to obtain the advantages of an electromagnetically-coupled transducer, as in, for example, a resolver, but avoids many of the difficulties inherent in the variable amplitude system. In the phase system the rotor coil, or moving coil, is magnetically linked with a number of primary coils, the net magnetic field producing a single output voltage which is constant in amplitude and has a phase angle (relative to the supply) directly proportional to the input displacement. Thus if the supply voltage is

$$v_p = V_p \sin \omega t \qquad \ldots\ldots(1)$$

the signal voltage required from a transducer energized by this supply is

$$v_s = V_s \sin(\omega t + \phi) \qquad \ldots\ldots(2)$$

where ϕ is a phase angle with a linear relationship to a displacement. If the displacement from a datum is x and the phase at $x = 0$ is ϕ_0, then over distances giving phase angles of less than 360° the desired equation for this form of transducer is

$$\phi = K_s x + \phi_0 \qquad \ldots\ldots(3)$$

or, in the rotary form,

$$\phi = \theta + \phi_0$$

where θ is an angle of rotation, and K_s is the change in phase per unit length and is usually a constant. Providing practical arrangements are made to account for the numbers of complete rotations of 360° then equation (3) may be extended over any angle. The phase ϕ_0 is usually a constant determined by the choice of the zero position of x, but in certain special applications ϕ_0 may be a variable. Assuming that no magnetic materials are to be included in the transducer design, the supply frequency will be higher than is usual for power systems, the preferred choice being in the range 16 kHz to 20 kHz.

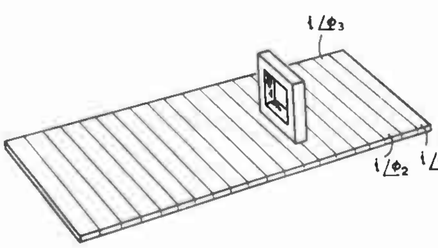


Fig. 1. An array of conductors forming a linear transducer.

The principles employed in the transducer may be illustrated by the linear transducer shown in Fig. 1. A coil is positioned over an array of equally-spaced parallel conductors and constrained to move along a path parallel to the surface of the array. The array is energized by alternating current, the amplitude of the current in each conductor being the same but the phase angle in successive conductors differing by a fixed amount. The voltage induced in the coil due to the flux linkage from all the conductors may be considered to be made up of components derived from individual conductors. Since the flux linkage will be greater from the conductors nearest to the coil, they will provide the main contribution to the net voltage and thus largely determine the phase of the voltage.

In Fig. 1, let the current in successive conductors differ in phase by angle α so that in phasor terms the voltage E induced in the coil can be written,

$$\begin{aligned} E = {} & |E_0|\angle 0 + (|E_1|\angle\alpha + |E_{-1}|\angle{-\alpha}) + \\ & + (|E_2|\angle 2\alpha + |E_{-2}|\angle{-2\alpha} + \ldots + \\ & + (|E_r|\angle r\alpha + |E_{-r}|\angle{-r\alpha}) \qquad \ldots\ldots(4) \end{aligned}$$

where $|E_r|\angle r\alpha$ is the component due to conductor r, and that due to conductor 0 being taken as the datum of phase. The net phase of E can be obtained by phasor addition of the components of (4). For a continuous array, and the coil over the '0' conductor it can be shown that, $|E_1| = |E_{-1}|$, $|E_2| = E_{-2}|$ etc., therefore from equation (4):

$$\begin{aligned} E = |E|\angle\phi &= |E_0|\angle 0 + 2|E_1| \cos\alpha \angle 0 + 2|E_2| \cos 2\alpha \angle 0 + \ldots \\ &= |E|\angle 0 \end{aligned}$$

That is, E has a net phase equal to that derived from the conductor directly below the coil. For other positions it can be shown by numerical calculation that the phase of the induced voltage varies uniformly as the coil is moved from one conductor to the next. Providing the array is energized with uniform steps of phase then equation (3) is satisfied. In a similar manner, where the array of conductors is arranged uniformly around a cylinder to form a stator, the phase of the induced voltage in a rotor coil is determined by the angular position of the coil with respect to the conductors.

3.2. *Applications*

The linear transducer is made in short sections in a similar manner to that used for printed circuit boards which can be connected in series to form a transducer of any length. Providing the positioning of the array of conductors is sufficiently accurate, the transducer can be produced to a positional accuracy of one thousandth of a pitch. A pitch (360° phase change) is typically 1 inch (2·54 cm) but could be much smaller.

The rotary transducers, which are available commercially, are for series connexion (current mode) or parallel connexion (voltage mode). In both cases the output is usually 360 electrical degrees per rotation, with the phase angle corresponding to the mechanical angle to within ± 10 minutes of arc for the voltage mode transducer, and two or three times better for the current mode transducer. In all cases the resolution is continuous.

The transducer may be used to give a feedback for a servo position control. A typical example is shown in Fig. 2, where the input signal phase angle is compared with the feedback signal phase angle in a phase discriminator, deriving a d.c. voltage proportional to the difference. When the two phases are equal, the output of the discriminator is zero, the null being defined to an error of less than 0·01 electrical degrees.

In many applications the servo requires a d.c. voltage proportional to the output velocity (tacho feedback).

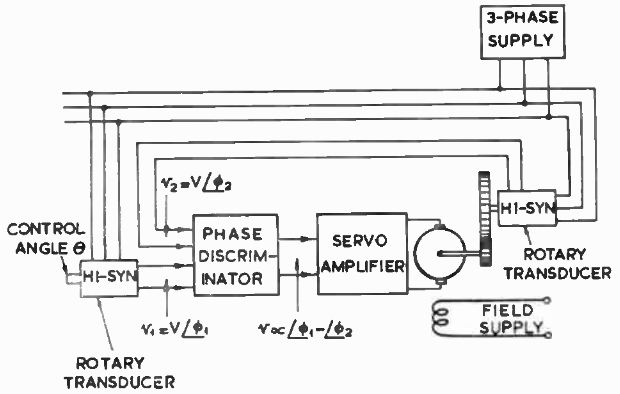


Fig. 2. A rotary transducer used in a servo position control.

This can be achieved in a simple manner, by converting the transducer output, thereby saving the inconvenience and cost of a tacho and mounting arrangements.

Units are commercially available for converting the output of a transducer into a digital number, conveniently a b.c.d. code of three decades. Similarly an electronic conversion can be made changing a digital number into a phase angle and providing an interface between the computer output and the servo input.

It has been shown that for the rotary transducer the phase of the output signal is given by

$$\phi = \theta + \phi_0$$

where the phase angle is with respect to the supply. It follows therefore that ϕ_0 is determined by the mechanical setting of the transducer and the supply phase. Consider two rotary transducers A and B. If the output of B is converted into a form suitable for the supply to A then the phase of A due to the combination is

$$\begin{aligned}\phi_A &= \theta_A + \phi_B \\ &= \theta_A + \theta_B + \phi_0\end{aligned}$$

Three useful results emerge from this, namely:

(1) The datum for a servo feedback can be changed by a fixed amount so that, for example, the contour of a program of inputs can be moved.

(2) Using two transducer arrays having different scale constants, e.g. 1 inch and 1·1 inch pitch, the difference or vernier effect can be obtained.

(3) A signal having a phase angle equal to the differential between two shafts rotations can be derived.

The ability to move the datum is particularly important in mechanical handling where the start or the finishing point of series of movements may be varied from time to time.

4. Programmed Control

From the range of program controllers available, the most accurate and most often used are those based on a stored program in digital form, in particular on punched-paper tape. Numerically-controlled machines have demonstrated the high accuracy and reliability that can be achieved. However, numerical programmers give point-to-point control so that where a smooth contour is required, and accuracy is not vital, they are inconvenient and more expensive than is necessary.

There are many forms of analogue programmer and the choice should be made for each application.

In mechanical handling, the following aspects are of major importance.

(1) Reliability. With the introduction of a larger number of automatic controls, the failure of any one of which disrupts production, reliability becomes of paramount importance.

(2) Power integrity. After a power failure, a program controller should restart, if required, at the correct new point in the program without any attention. A momentary power failure should not produce dangerous transients.

(3) The output signal from the programmer should be in a form suitable for use with the control system and displacement transducers.

Two programmers having a phase signal output will be described; both have very high inherent reliability. The first, which uses all solid-state components, can only produce a program consisting of a series of straight lines. The second, an electromechanical system, can produce an arbitrary contour.

4.1. *A Solid State Program Controller*

The system relies on the smooth change in phase obtained when a transition occurs between two added voltages having differing phase angles. Thus, if a new voltage $V/\underline{\phi_2}$ is added, in gradually increasing amounts, to the previous voltage, $V/\underline{\phi_1}$ which is reduced in a corresponding manner, then the resultant voltage has a relatively constant amplitude and a smoothly changing phase. This holds true for total phase changes up to 60 or 70 degrees. Figure 3 illustrates the process, ϕ_a, ϕ_b, and ϕ_c being typical intermediate positions.

Fig. 3. Diagram showing a change in phase.

Figure 4 shows three sections of a program controller. The sections are connected between three sections of a program controller. The sections are connected between the input control-voltage rail and the summing rail, the latter being the summing junction of the output amplifier. Assuming first that section 1 is 'active' and all higher sections 'inactive' then the output of A1 is high and positive, transistor TR1 is open, Z_1 is high, and V_1 is proportional to $V/\underline{\phi_1}$. Now, as the control voltage goes

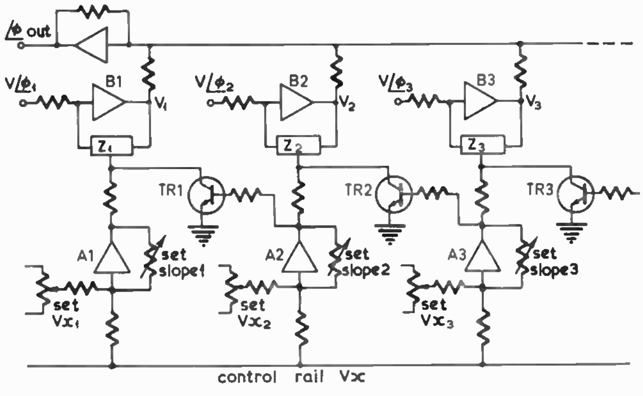


Fig. 4. A solid state program controller.

progressively negative a point is reached, determined by the setting of Vx_2, where the output of A2 goes positive. This causes TR1 to pass current, progressively reducing Z_1 and V_1. At the same time Z_2 is increased, increasing V_2 until as V_1 becomes zero, V_2 becomes $V/\underline{\phi_2}$. The slope of the transition is set by the resistance across A2.

As the control voltage continues to go negative, section 3 is made active, 2 becomes inactive, and 1 remains inactive. In this way a continuous programme of 'slopes' and 'flats' can be set.

Although the detail in the contour is limited in this type of programmer, it has the considerable merit that it can be set *in situ*, or adjusted during a working operation, thereby, gaining optimum performance.

Figure 5 shows a typical program using three sections.

4.2. *An Electromagnetic Program Controller*

It is often a requirement that a series of movements should follow a continuously curved path, and the limited contours of 4.1 are not acceptable. Contemporary systems meeting this requirement take various forms, for example, cam followers, photo-electric curve followers etc., the common feature being, in the majority of cases, that a sensor is made to follow the prescribed curve. It is then necessary to obtain the output from a further transducer attached to the follower mechanism.

To obtain a higher order of reliability and using phase as the form of signal, a system was devised which relies on the electromagnetic coupling between a wire and series of coils.

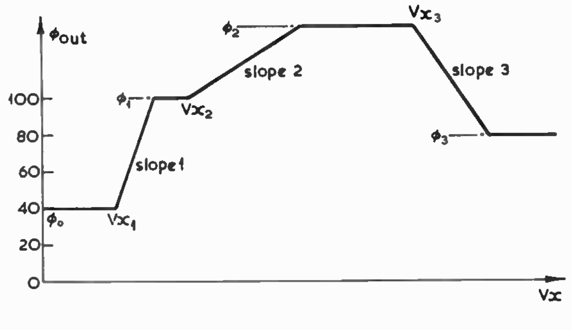


Fig. 5. Typical program for the solid state program controller.

In Section 3 it was shown that an array of conductors and a single coil would produce a signal of the required form. An alternative arrangement, and equally effective, is an array of coils positioned over a single wire. In this case the output from each coil (starting from the end coil) is progressively modified in phase by equal amounts by added phase shift networks, until the output from the final coil requires a phase shift of 360° (i.e. zero phase shift). All the signals are then passed to a summing amplifier so that the phase of the output signal is determined by the relative position of the wire under the array of coils. (See Fig. 6.)

In a practical version the conducting wire would take the form of a thin conducting strip, printed on a long strip of flexible printed circuit. The curved path of the signal required would be derived by the curved path of the printed circuit. The printed circuit, normally stored in a cassette, is drawn across the pick-up coils, in a similar manner to that used for tape recorders, and as the tape progresses, a program of signals is issued.

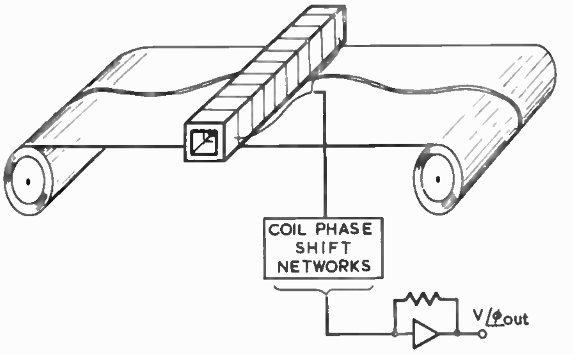


Fig. 6. A continuously curved path program controller.

4.3. *Datum Shift*

One difficulty frequently encountered with programmed control of mechanical assembly is that the position of each workpiece when it arrives at an assembly station varies with respect to the local datum. For example, in car assembly, the body of the car arrives at an assembly station on a ground supported truck, and the car sides by overhead conveyor. In contemporary methods the car sides, as they approach, are held in a jig which is made, with considerable force, to align itself with the truck by a mechanical interlock.

It will be appreciated that position errors of the car body on the truck and the sides in the jig and the crudeness of the interlock produce overall errors limiting the application of this method.

To proceed with more sophisticated programmed assembly it is necessary to have sensors to determine the actual positions of the parts to be assembled. The sensors should then modify (update) the relevant part of the program. Methods to achieve this are currently being studied, and one method, depending on a small magnet attached to the car and parts, has been tried.

5. The Automatic Warehouse

For automatic point-to-point transportation, a continuous detailed control of the vehicles is unnecessary.

It is sufficient to exercise continuous control within a short range of each specified point and to permit a free movement between points. During free movement the vehicle must retain a sense of direction, avoid obstacles (if not constrained to tracks) and sense the stopping point sufficiently in advance to permit a controlled deceleration.

In the following sections, an extension of the phase angle transducer system will be described as applied to an automatic warehouse, although the system could be used for any point-to-point transport.

The vehicle, called a stacker, moves along an aisle between two high-density storage structures. A load may be retrieved, from the left or right, and at any height, by a mechanism carried on the stacker mast. The mast is usually vertical, so that, if the stanchions of the storage structure are also vertical, a location system derived from the ground position will position the stacker correctly for retrieving at any height. However, it often imposes a costly premium or an unwanted design condition, to specify a vertical accuracy for the structure. To overcome this difficulty one known method employs a reflective ribbon attached to the stanchion and a photoelectric detector carried by the stacker is used for final location. The particular stanchion required is located by an additional system; usually a numerical, photoelectric detection system.

To obtain a better order of reliability than obtainable by photo-electric methods, which generally require a clean environment, a system based on the electromagnetic detection of a wire was proposed. The system, although described for horizontal positioning, could be applied for vertical or other axes.

5.1. *The Wire Location System*

A single wire for each horizontal location, i.e. each stanchion, is passed from a controller along the ground, up the appropriate stanchion and finally earthed to the metal structure. In this way the horizontal position of every opening is a predetermined distance from a wire.

A pair of coils, carried by the retriever platform as shown in Fig. 7 can be positioned so that when the coils are equal distances from a wire, the retriever is in the correct position for a withdrawal. The wire carries a.c. which, in this position, induces an equal voltage in each coil. A movement from the centre position produces an increase in voltage in one coil and a reduction in the other. The voltages are added together via a phase-shift network so that, in a manner similar to that described in Section 4, the output phase is changed by the relative position of the coils with respect to the wire, the phase of the centre position being determined by the phase of the current in the wire. It may be assumed for convenience that the 'centre phase' is zero, and that a small displacement to the right produces a positive change of phase, and a displacement to the left a negative. The stacker motor control ensures that a phase error signal sends the stacker in the correct direction so that in this case a positive phase produces a drive to the left.

The maximum phase change that can occur (when one coil is above the wire), is limited to a value set by the phase shift network. Let this be ϕ_c.

If the coils are moved far away from the wire the signal amplitude drops to a low value giving indeterminate results. Before this threshold value is reached a 'memory circuit' is initiated which introduces an input voltage having a phase of $+2\phi_c$ or $-2\phi_c$, maintaining full motor power in the correct direction between wires.

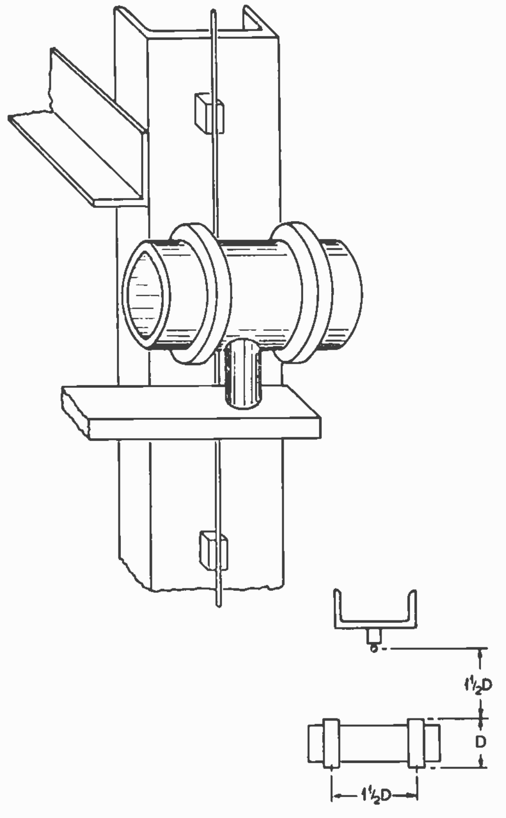


Fig. 7. The wire location system.

5.2. *The Array Control*

The controller, as determined by an operator, a computer, or tape, selects the particular wire, corresponding to a particular position required. The selection initiates an electronic circuit which changes the phase of the current in the chosen wire into zero phase, at the same time modifying all the circuits for wires to the right of the chosen wire to give a positive phase equal to $2\phi_c$ and modifying all the circuits for wires to the left to give $-2\phi_c$. With this arrangement, no matter where the stacker starts, on reaching a wire, the stacker controller will receive a signal of $2\phi_c$ (or $-2\phi_c$), setting the 'memory circuit' and the motor power for full speed in the correct direction. The stacker, passing all unselected wires, will continue to travel in the correct direction, until the chosen wire is reached, where it will stop.

5.3. *Deceleration Control*

The distance between the coils provides a zone of 'control proportional to displacement', but the stacker, approaching this zone at full speed, requires considerably more than this distance to slow down. A deceleration system is therefore essential.

A signal can be readily obtained by an attachment to the stacker near the ground carrying two coils, one in front and one behind the stacker. As the leading coil passes over a wire carrying current of zero phase, a 'stop' signal is detected and the deceleration initiated.

In order to obtain a low deceleration time, and therefore higher efficiency, it is desirable to make the rate of deceleration proportional to the initial velocity (i.e. a constant stopping distance). A close approximation to this was achieved in a laboratory model by using the deceleration signal to change the form of the control feedback. During normal running the control employed acceleration feedback enabling the motor to reach full speed. After the deceleration signal the control changed to velocity feedback reducing the motor speed in a controlled transition to produce the required deceleration characteristic.

5.4. *Coil Spacing and Accuracy*

The location accuracy depends upon the size and spacing of the location coils and the closeness to the wire. If D is the coil diameter a typical arrangement would be, a coil spacing of $1{\cdot}5\,D$ with each coil a distance of $1{\cdot}5\,D$ from the wire. With this spacing, a location accuracy of 1/30 of the coil diameter is possible. For example, a 3 in (7·6 cm) coil with 4·5 in (8·9 cm) wire clearance and a control zone of 4·5 in would achieve a location accuracy of $\pm 1/10$ in (2·54 mm).

It should be noted that the stacker positional accuracy also depends on other additional factors, for example, backlash in gears, structural resilience, etc.

6. Conclusions

A transducer and other devices based on electromagnetic coupling producing a signal where the phase angle is proportional to displacement has been described and shown to have many applications in connexion with mechanical handling.

In developing the automatic systems described particular emphasis was placed on the reliability requirement which, it is suggested, is a key factor in further progress. Further progress also depends on an overall consideration of the product, the manufacturing technique, resources, and a planned changeover from manned to automatic methods.

7. Acknowledgments

The author wishes to acknowledge the facilities provided by the University of Salford and Mr. D. McEwan for his help and encouragement, also Mr. K. Wilkinson of Geo. W. King Ltd., who initiated most of the work.

Manuscript received by the Institution on 4th May 1971. (*Paper No. 1428/IC57.*)

U.D.C. 621.376.2

Double-Tuned Modulator Calculations Using an Improved Diode Model

K. J. GLOVER, B.Sc., M.Sc., C.Eng., M.I.E.E.*

J. G. GARDINER,
Ph.D., C.Eng., M.I.E.E., M.I.E.R.E.*

and

Professor D. P. HOWSON,
D.Sc., C.Eng., F.I.E.E., F.I.E.R.E.*

SUMMARY

Linear time-varying analysis for single-balanced diode mixers and modulators is developed to incorporate a non-linear conductance model of the non-linear elements employing the exponential *i/v* characteristic and series resistance of a typical Schottky-barrier diode. It is shown that over a significant range of local-oscillator drive conditions calculations based on the much simpler bilinear diode model can be expected to yield predictions of loss and matching conditions which closely approach those obtained from the more sophisticated calculations. However, in operating conditions involving the use of d.c. bias voltages approaching the peak local-oscillator voltage, as is necessary for the achievement of low loss in certain double-tuned circuits, significant departures occur resulting in failure of the 'duality' relationships existing between series and shunt configurations. The series circuit is shown to be theoretically capable of superior loss performance to the shunt, but under terminating conditions which will be difficult to realize in practice.

* *Postgraduate School of Electrical and Electronic Engineering, University of Bradford, Bradford 7.*

List of Symbols

A_L	mixer conversion loss in dB
$g(t)$	exponential diode conductance variation
I	current delivered by local oscillator generator
I_D	exponential diode junction current
I_S	exponential diode reverse leakage current
i_0, i_{+1} etc.	components of signal circuit current at signal, upper-sideband frequencies, etc.
K	Boltzmann's constant
q	electronic charge
r_b	bilinear diode reverse resistance
r_f	bilinear diode forward resistance
R_D	exponential diode series resistance
R_R	exponential diode effective reverse resistance
R_P	local oscillator source resistance
$r(t)$	exponential diode resistance variation
$r'(t)$	bilinear diode resistance variation
s	ratio mark/mark-plus-space, bilinear diode resistance variation
T	temperature °K
V_b	d.c. bias voltage
V_J	exponential diode junction voltage
$\hat{V}_P$	peak local-oscillator voltage
$\hat{V}_q$	peak signal voltage
ω_p	local oscillator frequency
ω_q	signal frequency

1. Introduction

In common with conventional linear circuits of all types frequency-changing devices are being realized increasingly in microelectronic form for operation over many frequency ranges from h.f. to microwave. Whilst many of these circuits appear simple in design and construction, a wide range of parameters, particularly of the characteristics of the non-linear elements, frequency selective properties of the terminations, etc., exercise significant influence over conversion loss, optimum terminating impedance and matching conditions. In particular, recent advances in Schottky-barrier diode manufacturing techniques have resulted in renewed interest in the extent to which the parameters of these devices represent a limitation to the further reduction of conversion loss in diode mixers and modulators and to the implementation of circuit techniques at microwave frequencies which have for many years been known to give low-loss circuits at low frequencies.[1] It has for some time been appreciated that losses approaching zero decibels are theoretically attainable in mixers with input signal and load terminations tuned in such a way that dissipation is permitted only at the wanted modulation sideband (in the load) and input signal (in the source circuit).[2] It is further known that dissipation of power by unwanted products has a significant influence on the circuit loss to wanted sidebands, in particular the 'broadband' microwave mixer, which permits dissipation of image power in the input circuit, has a minimum loss of 3 dB.[3]

However, practical attempts to realize either zero-loss or ideal 'broadband' mixers seldom succeed in producing circuits whose losses approach nearer than 2 dB to the theoretical minimum, and it has become a matter of considerable interest to isolate the extent to which measured departures from the ideally-obtainable losses can be attributed to imperfections in the terminating circuitry, resulting in unwanted dissipation, as opposed to imperfections in the diodes associated with the shape of the device characteristic, series resistance, etc.

In approaching this problem from a theoretical point of view two major techniques are available:

(1) Time-domain analysis for general non-linear circuits.

(2) Frequency domain analysis incorporating a linear periodically-varying-parameter representation of the non-linear element in which the coefficients of the Fourier series representing the parameter variations are derived from interaction of the local-oscillator signal with the exponential diode i/v characteristic and series resistance.

The first technique, whilst applicable in principle, has significant disadvantages:

(a) Since results are obtained by spectral analysis of an output waveform constructed by iterative computational procedures over a large number of discrete time intervals, the signal and local-oscillator frequencies are inevitably constrained to be harmonically related so that the output waveform periodicity can be restricted to a realistic figure. In many modulator applications this restriction hinders interpretation of the spectral analysis since modulation products, intermodulation products and signal harmonics are superimposed at many frequencies of interest and isolation of the contributions to a given spectral line from a particular side-band under investigation becomes impossible.

(b) Formulation of time-domain analysis programs to achieve great generality of application results in their giving little insight as to the influence of particular circuit parameters on performance characteristics so that 'trial-and-error' variations of many parameters are necessary in order to build up some picture of the operating mechanism. This requires in many applications an uneconomic amount of computation.

At the other extreme, analyses to date using linear time-varying techniques have assumed very elementary diode models—often a bilinear approximation—in order to produce simple analytical expressions for mixer and modulator conversion loss, input impedance, etc.[4] However, as will be demonstrated in the present paper, the incorporation of a suitable diode model in linear time-varying analysis can result in a very satisfactory compromise between the accuracy of circuit modelling obtainable with time-domain analysis and the ease and rapidity with which results can be calculated using conventional linear time-varying methods. This permits description of the diode in terms of its exponential i/v characteristic with associated series resistance but involves acceptance of the assumption common to all linear time-varying analysis with regard to selectivity of terminations; namely that these present resistive terminations at their resonant frequencies and either zero or infinite impedance elsewhere.

2. Small-signal Resistance and Conductance Variations for Pumped Diodes

Linear time-varying analysis assumes, by definition, that the conductance (or resistance) variation of the non-linear element is controlled entirely by the local-oscillator, i.e. that the slope conductance or resistance of the diode characteristic is effectively linear over the signal voltage excursion. Thus, from an analytical point of view the influence of the diode characteristic shape appears in the relative magnitudes of the coefficients in the Fourier series representing the periodic variation of the device conductance (or resistance) over the local-oscillator cycle, i.e.

$$g(t) = g_0 + \sum_{n=1}^{\infty} g_n \cos n\omega_p t \qquad \ldots\ldots(1)$$

$$r(t) = r_0 + \sum_{n=1}^{\infty} r_n \cos n\omega_p t \qquad \ldots\ldots(2)$$

A number of authors have derived values for the coefficients of the above expressions by considering the exponential diode characteristic[5,6] but in order to derive these analytically have assumed that the local oscillator supply establishes a sinusoidal voltage across the junction itself, subsequently incorporating the diode loss resistance in the small-signal circuit analysis. In the present work this restriction is removed by using time-domain analysis to predict the time variation of $r(t)$ and $g(t)$ to permit greater modelling accuracy and flexibility.

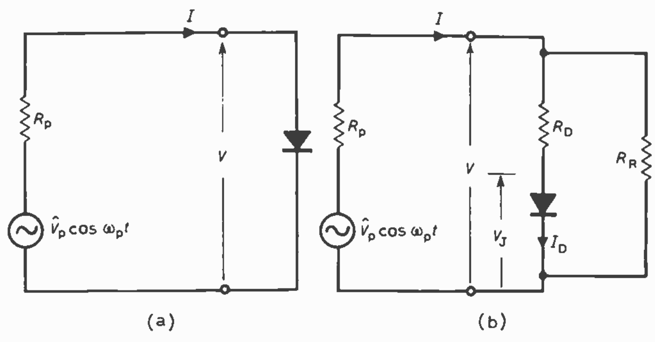
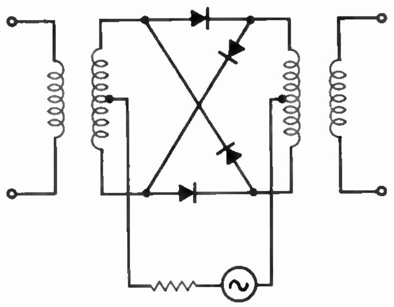
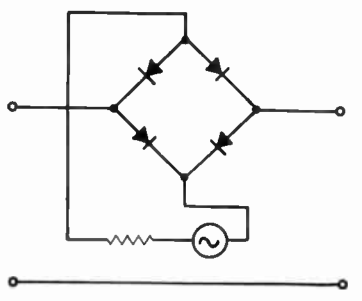


Fig. 1. (a) Elementary untuned local oscillator circuit; (b) Diode model for computation

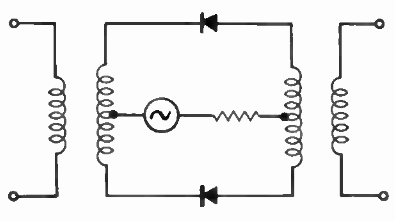
Referring to the diagrams of Fig. 1(a) and (b) the device is seen to be represented by its non-reactive characteristics and to be embedded in a simple untuned local-oscillator circuit. (The assumption of a broadband local-oscillator circuit is well justified for the majority of practical diode mixers and modulators operating up to v.h.f. since such circuits almost invariably employ some form of balanced local-oscillator circuit such as those shown in Fig. 2.)



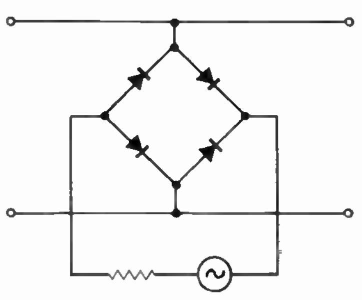
(a) Ring



(b) Series



(c) Half-ring



(d) Shunt

Fig. 2. Typical balanced local-oscillator circuits.

Therefore, if the applied local-oscillator voltage V_p is assumed to include d.c. bias, V_b, and to be applied from a generator of resistance R then

$$V_p = \hat{V}_p \cos \omega_p t + V_b \qquad \text{......(3)}$$

and assuming the exponential relationship between junction current and voltage, i.e.

$$I_D = I_S[\exp(qV_J/nKT) - 1] \qquad \text{......(4)}$$

where I_D and V_J are as defined in Fig. 1(b) and I_S is the reverse saturation current. R_D is included to simulate the forward resistance of the diode due to bulk resistance in the semiconductor, etc. and R_R represents the finite reverse resistance which the diode achieves under reverse bias. The latter was felt to be necessary in order to accommodate effects of such phenomena as charge storage, etc. which, whilst minimized in the Schottky-barrier device, nonetheless result in a much smaller effective reverse resistance under high-frequency switching conditions than that suggested by manufacturers' data for the d.c. characteristic of (4).

Computation proceeds in the usual manner—values of V_p being taken at various $\omega_p t$ and V_J computed iteratively using standard techniques. The diode conductance thus presented by the diode to the small-signal circuit is therefore given by

$$g(t) = \frac{dI}{dV} \qquad \text{......(5)}$$

The diode incremental resistance is calculated similarly as

$$r(t) = \frac{dV}{dI} \qquad \text{......(6)}$$

and repetitive conductance and resistance variations of the form of Fig. 3 are readily constructed. Representation in the form of equations (1) and (2) requires only application of standard fast Fourier transform procedures.

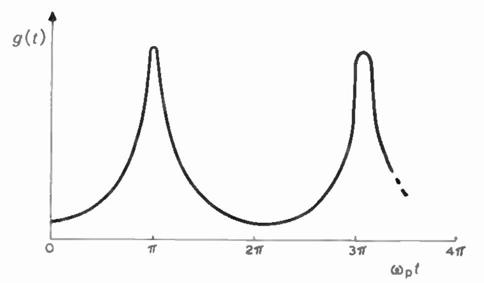
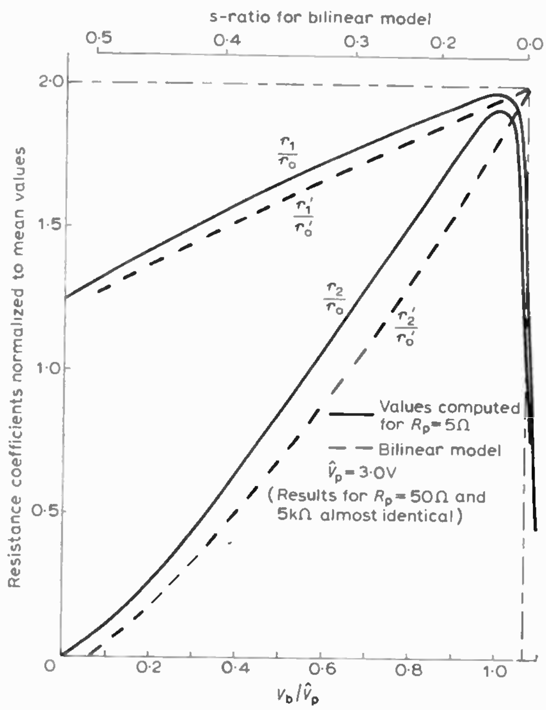


Fig. 3. Typical small-signal conductance variation for pumped diodes.

Thus it is apparent that not only can the influence of the shape of the diode d.c. characteristic be readily evaluated but also the effects of such circuit considerations as local-oscillator e.m.f. source resistance and d.c. bias become accessible.

We are now in a position to examine the predictions of the exponential law diode model in comparison with those derived from the more usual bilinear representation of the device. Direct comparison is difficult because of the two properties of the bilinear diode model which are not shared by the exponential law representation.

(1) The bilinear device characteristic is discontinuous so that an infinitesimally small change in applied local-oscillator voltage at the point of discontinuity results in a complete and instantaneous change in the diode state from completely turned 'off' to completely turned 'on' or vice versa. In other words the linear time-varying resistance/conductance function produced by the local oscillator is entirely independent of the magnitude of the local-oscillator voltage.



Diode parameters

$\frac{q}{nKT} = 40\ \text{V}^{-1}$

$R_D = 11\Omega$

$I_s = 1 \times 10^{-9}$A

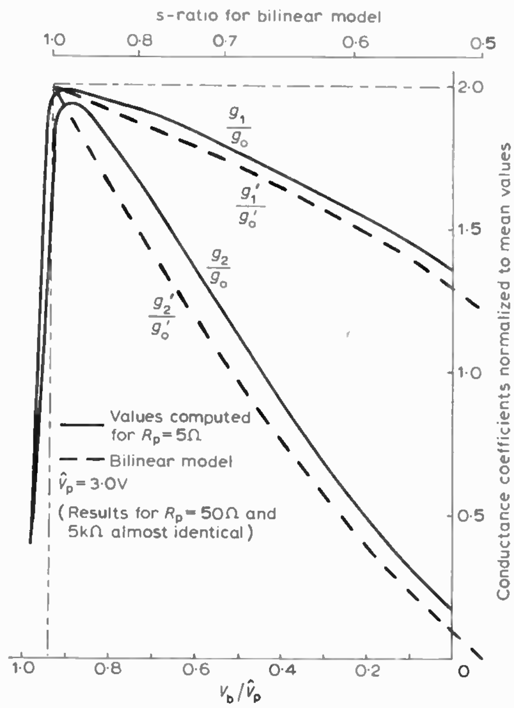


Fig. 4. Comparison of resistance coefficients predicted by bilinear and exponential diode models.

Fig. 5. Comparison of conductance coefficients predicted by bilinear and exponential diode models.

(2) As a consequence of (1) the resistance/conductance function is always a rectangular waveform of zero rise and fall times with a mark/space ratio determined by the voltage assumed necessary to turn the diode on (the point of discontinuity in the characteristic) relative to the local oscillator and d.c. applied voltages. This means that it is possible to envisage the use of a variable d.c. bias to adjust the mark/space ratio from some value approaching unity through 0·5 to some value approaching zero.

Plainly, the exponential-law device model cannot share these properties since its characteristic is continuous; over a range of local-oscillator voltages from a few millivolts to several volts the resistance/conductance function changes from being virtually constant with a very small almost sinusoidal fluctuation to a rectangular waveform approaching that predicted by the bilinear model with a range of intermediate conditions depending on the exact values of the exponent in the characteristic equation, the device series resistance relative to the local-oscillator source resistance etc. Moreover, variation of the d.c. bias applied with the local oscillator cannot be expected to produce the changes in mark/space ratio achievable in the bilinear case since the resistance function is no longer of rectangular form with zero rise and fall times—100 mV or so is required to change the effective state of the device.

However, from a practical point of view the parameters available for adjustment in an experimental circuit are local-oscillator e.m.f. and source resistance and the level of d.c. bias maintained at the diode (either by direct application or by rectification of local-oscillator current), and these considerations have formed the basis for the comparisons shown in Figs. 4 and 5. Here it is assumed that a typical local-oscillator drive situation has been established, 2V r.m.s. e.m.f. and 50 Ω source resistance, and that the d.c. bias voltage is variable. At the drive e.m.f. assumed it is to be expected that significant departures from the predictions of the bilinear model will emerge as attempts are made to produce effective mark/space ratios which tend to unity or zero which, as will be seen in the next section, are required for low loss operation of two particularly useful modulator circuits. Figure 4 therefore shows the variation of $r(t)$ as the diode is biased to increase the conduction time per local-oscillator period. At each value of $V_b/\hat{V}_p$ we make a comparison with the bilinear model by assuming the change of state to occur at 250 mV approximately so that when $V_b + \hat{V}_p \cos \omega_p t > 250$ mV the device conducts. This enables us to define a mark/mark-plus-space ratio s from which the coefficients r_0', r_1' etc. are readily deduced from the expression:[2]

$$r'(t) = r_f + s(r_b - r_f) + \frac{2}{\pi}(r_b - r_f)\sum_{n=1}^{\infty} \frac{\sin n\pi s}{n} \cos n\omega_p t \qquad \ldots\ldots(7)$$

The values of s which correspond to various bias levels are shown in Fig. 4.

Figure 5 shows a similar calculation for $g(t)$ as the bias is adjusted to reduce conduction time and again the effective value of s for the bilinear model is indicated on the normalized bias axis.

In both graphs the coefficients r_1, g_1 and r_2 and g_2 are normalized with respect to the relevant mean resistance or conductance value.

The interesting features of these graphs are:

(a) A close correspondence exists between the predictions of the exponential and bilinear models over a wide range of bias conditions.

(b) In both cases the normalized values of the first harmonic coefficients r_1/r_0 and g_1/g_0 approach closely to the optimum value of 2·0 indicating, as will be apparent from the next Section, that low loss is attainable.

Thus it is apparent that the bilinear approximation gives an accurate representation of the device operation for values of bias voltage resulting in mark/space ratios of 0·1 to 0·9 but for more extreme conditions, such as are required to realize the low losses reported in reference 2, representation of the device in terms of its exponential characteristic is advisable.

3. Performance of Typical Double-tuned Mixers

3.1. *Loss and Termination Requirements in Terms of $r(t)$ and $g(t)$*

The circuits to be considered are those shown in Fig. 6(a) and (b), and are chosen because their conversion loss can ideally be made to approach zero dB provided satisfactory circuit conditions and diode properties can be obtained simultaneously.[2] It is convenient to analyse the shunt circuit by way of illustration: referring to the equivalent circuit of Fig. 6 (c) we observe that the diode circuit i_d is constrained to contain only two components of signal and wanted modulation-product frequencies so that the circuit equation becomes

$$\hat{V}_q \cos \omega_q t = [i_0 \cos \omega_q t + i_{1+} \cos (\omega_q + \omega_p)t] \times$$
$$\times \left[R_q + r_0 + \sum_{n=1}^{\infty} r_n \cos n\omega_p t\right] \qquad \ldots\ldots(8)$$

which can be written at each frequency involved separately and solved for the sideband current i_{1+} to give

$$i_{1+} = \frac{2r_1 \hat{V}_q}{r_1^2 - 4(R_q + r_0)^2} \qquad \ldots\ldots(9)$$

which, using the conventional power definition of conversion loss A_L, gives

$$A_L = 10 \log \frac{[r_0 + \sqrt{r_0^2 - \frac{1}{4}r_1^2}]^2}{\frac{1}{4}r_1^2} \qquad \ldots\ldots(10)$$

under optimum terminating conditions of

$$R_q = \sqrt{r_0^2 - \tfrac{1}{4}r_1^2}. \qquad \ldots\ldots(11)$$

It is apparent from inspection of (10) and (11) that if the conditions $r_1 = 2r_0$ can be approached then A_L tends to zero dB under terminating conditions which approach short circuit. This condition is approached in principle by an ideal bilinear diode of arbitrarily high switching quality (i.e. the ratio of reverse to forward resistance tends to an infinitely large value) when this is arranged to conduct for almost the entire period of the local-oscillator waveform turning off only at the peak negative value of $\hat{V}_p$. For the bilinear diode this is attainable in principle with sinusoidal drive by choosing $V_b \simeq \hat{V}_p$ since the 'ideal' device attains its full reverse resistance value instantaneously and for an indefinitely small reverse bias voltage.

It should be borne in mind, however, that zero-loss conditions correspond to terminating impedances—open circuit in the series case, short circuit in the shunt case—which are difficult to match to in practical circuit arrangements. As a result some relaxation of the drive conditions to obtain a compromise between loss and matching is necessary in practice.

Plainly, the limiting conditions necessary for zero loss in this circuit make very severe demands on the circuit and diode properties, so that the use of the best available diode model is necessary to obtain satisfactory evaluation of performance.

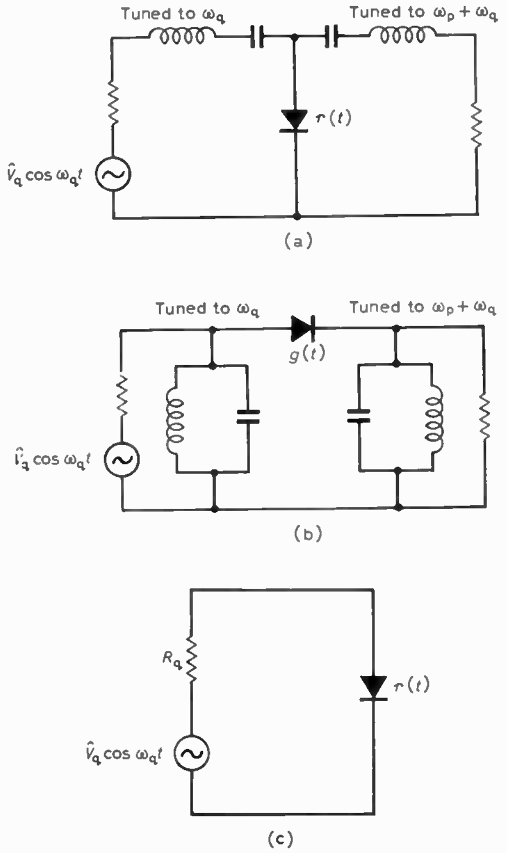


Fig. 6. Double-tuned mixers giving zero loss under 'impulse' local oscillator drive conditions: (a) shunt, (b) series, (c) equivalent circuit for (a). (R_q assumed infinite at all frequencies except ω_q and $\omega_p + \omega_q$.)

It is also evident that the ideal bilinear diode described in terms of $r(t)$ could also be described in terms of $g(t)$ and possess equally ideal conductance properties, so that an entirely dual set of results can hold for the series circuit of Fig. 6(b) in which zero loss is approached when the diode is turned off for almost all the time and the optimum terminating resistances tend to open circuits. However, an exponential device does not become effectively its own dual when transferred from a series to a shunt mixer so that a different performance is to be expected in practice when the same diode is used in the two different circuits.

3.2. *Influence of Diode and Circuit Parameters—Optimization*

3.2.1. Choice of bias for minimum loss

Figure 7 illustrates several of the above points. The performances of both shunt and series mixers are shown assuming that in each circuit the time-varying element is provided by a bridge arrangement of four diodes as indicated in Fig. 2(b), so that analysis of the circuit of Fig. 1(a) gives the required Fourier coefficients for $r(t)$ and $g(t)$. (The other balanced arrangements of Fig. 2 could be analysed with only minor modifications of the argument presented in Section 2.)

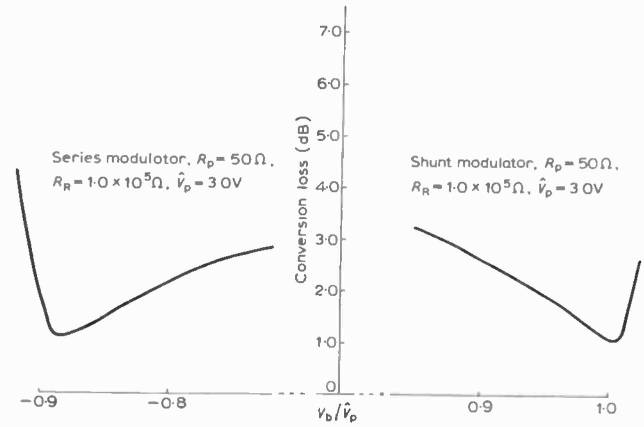


Fig. 7. Variation of conversion loss with bias for series and shunt modulators.

For the purposes of Fig. 7 the local-oscillator voltage was assumed to be 3 V e.m.f. peak supplied by a source of output resistance 50 Ω. As in the case of Figs. 4 and 5 virtually no difference is appreciable at this drive level between output resistances of 5 Ω and 5 kΩ. The diode parameters were taken from manufacturers' data for typical Schottky-barrier diodes (Hewlett-Packard type 2800) and were:

$$R_D = 11\ \Omega.\ \frac{q}{nKT} = 40\ \text{V}^{-1}$$

and

$$I_S = 8 \times 10^{-9}\ \text{A}.$$

A reverse resistance, R_R, of $1 \times 10^5\ \Omega$ was assumed and the major feature of interest is that, even with a diode possessing a relatively poor quality factor, losses of the order of 1·0 dB are obtainable in both series and shunt configurations, provided suitable matching is achieved in the terminations, and circuit losses, including dissipation by unwanted products, are negligible.

3.2.2. Influence of drive voltage and diode reverse resistance

Figures 8 and 9 show the results of plotting conversion loss against $V_b/\hat{V}_p$ for various values of $\hat{V}_p$ and a range of values of R_R from $10^4\ \Omega$ to $10^6\ \Omega$. It is of interest to note here that the shunt mixer of Fig. 9 is significantly less sensitive to variations in both $\hat{V}_p$ and R_R than the series mixer of Fig. 8, but that with high quality diodes ($R_R = 1 \times 10^6\ \Omega$) a significantly better loss performance

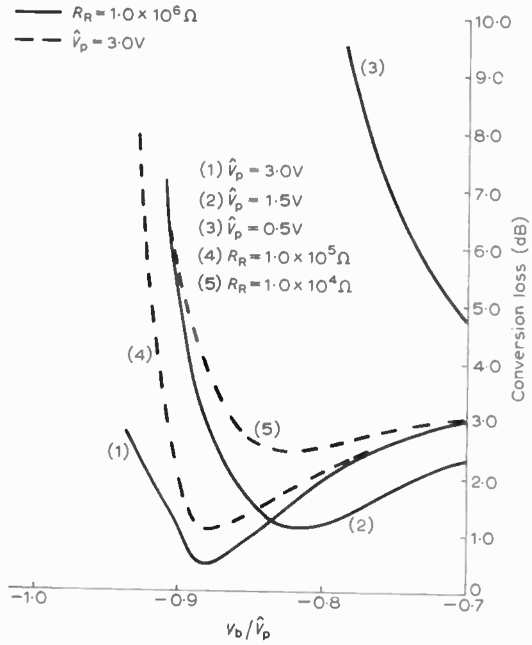


Fig. 8. Variation of conversion loss with local-oscillator source voltage and diode reverse resistance (series modulator).

is possible in the series case than the shunt, provided the necessary high resistance terminations can be realized (see Fig. 10).

3.2.3 Influence of local-oscillator source resistance

This aspect of performance is illustrated for both series and shunt mixer configurations in Fig. 10 and 11 respectively. In both cases $|V_b/\hat{V}_p|$ is maintained at 0·9 which results (from Fig. 7) in different values of

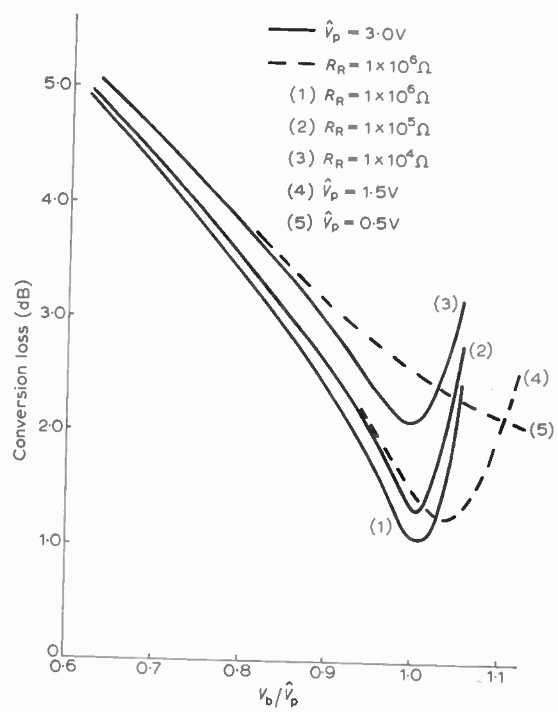


Fig. 9. Variation of conversion loss with local oscillator source voltage and diode reverse resistance (shunt modulator).

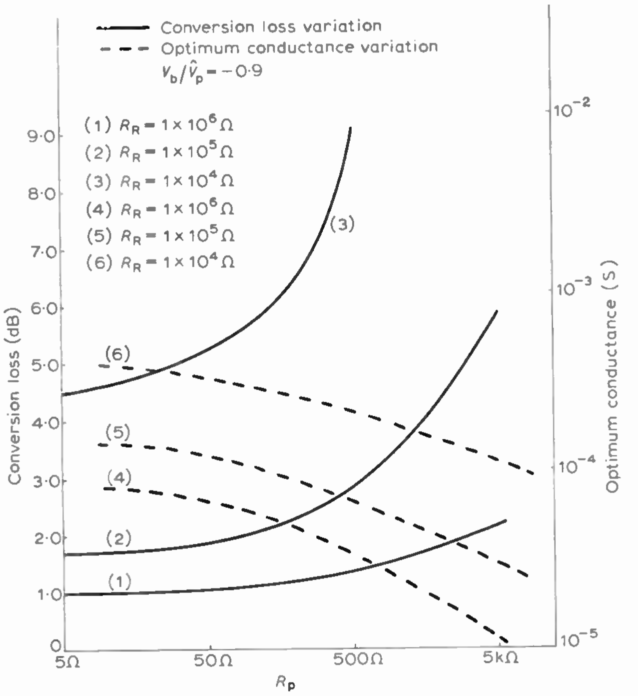


Fig. 10. Variation of conversion loss and optimum terminating conductance with local-oscillator source resistance (series modulator).

conversion loss, but it is again apparent that variations in this parameter have significantly greater influence on the series circuit than on the shunt, particularly for diodes with poor quality factors, i.e. reverse resistances $10^4\ \Omega$ or less. From a practical point of view this order of resistance ($10^4\ \Omega$) is probably the most realistic value for devices working at switching frequencies above 100 MHz. In these circumstances, it is apparent that a local-

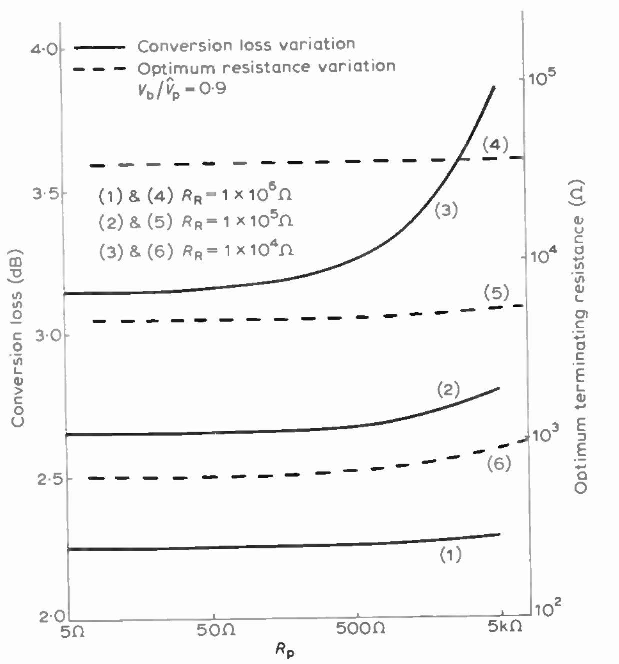


Fig. 11. Variation of conversion loss and optimum terminating resistance local-oscillator source resistance (shunt modulator).

oscillator source approaching an ideal voltage source is desirable.

3.2.4. Optimum terminating conditions

These are shown in Fig. 12 in conjunction with the remaining curves of Figs. 10 and 11. Again, it is seen that the shunt modulator has the most satisfactory performance with optimum terminating resistances of the order of a few hundred ohms around the bias condition for minimum loss.

4. Conclusions

The relative extent to which circuit imperfections on the one hand and limitations imposed by the characteristics of practical Schottky-barrier diodes on the other prevent the achievement of theoretically obtainable mixer conversion losses of less than 1 dB has long been a matter of interest and conjecture. It has been the purpose of the

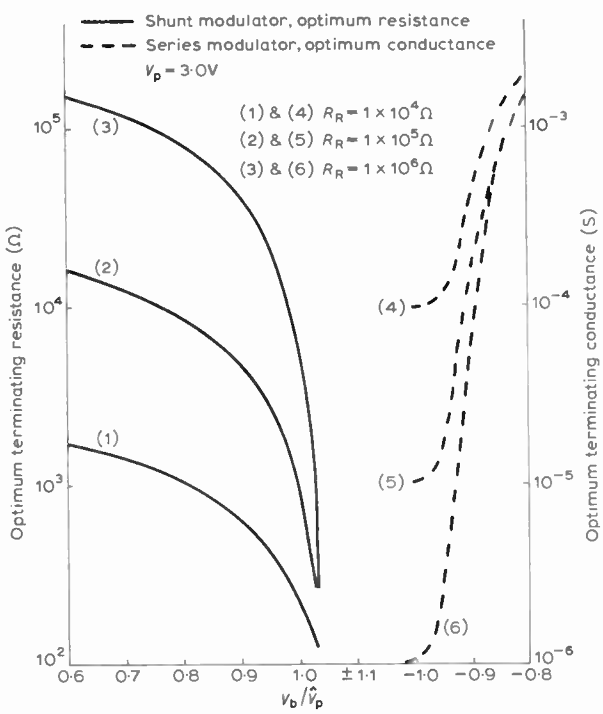


Fig. 12. Variation of optimum terminating conditions with bias (shunt and series).

present paper to clarify the influence of the diode characteristic and series resistance in conjunction with the local oscillator circuit on the loss of the series and shunt narrow-band mixers assuming that the input and output circuitry is ideally selective and loss free.

The calculations show that whilst the form of the diode characteristic does not appear to represent a limitation to the attainment of low loss, provided diodes of acceptable high switching quality can be obtained, from a practical point of view, the circuit conditions under which losses of less than 1 dB are obtainable must prove difficult to realize—very high Q resonant circuits or filters at unfavourable impedance levels (very high impedance in the series case) and so on. Also, as is to be expected of any mixer approaching ideal 'zero loss' operation sensitivity to variations in such circuit parameters as local-oscillator drive level, etc., becomes a significant problem, so that it is reasonable to conclude that operation in the region of 1·5–2·0 dB loss represents the most satisfactory compromise with existing devices. It is also possibly pertinent to note that, since the exponential device characteristic results in particularly favourable parameter variations for the series circuit that the advent of a device possessing such a characteristic, but with extremely small values of R_D, together with a value of R_R of a few hundred ohms, would permit the realization of very low loss mixers.

5. References

1. Barber, M. R., 'Noise figure and conversion loss of the Schottky-barrier mixer diode', *I.E.E.E. Trans. on Microwave Theory and Techniques*, MTT-15, No. 11, pp. 629–635, November 1967.
2. Howson, D. P., 'Single-balanced rectifier modulators', *Proc. Instn Elect. Engrs*, 109C, pp. 357–364, 1962.
3. Howson, D. P., 'Minimum loss of a general broadband mixer', *Electronics Letters*, 5, No. 6, pp. 123–5, 20th March 1969.
4. Tucker, D. G., 'The input impedance of rectifier modulators', *Proc. Instn Elect. Engrs*, 107B, pp. 273–281, 1960.
5. Torrey, H. C. and Whitmer, C. A., 'Crystal Rectifiers', Radiation Laboratory Series, Vol. 15. (McGraw-Hill, New York, 1948).
6. Liechti, C. A., 'Down-converters using Schottky-barrier diodes', *I.E.E.E. Trans. on Electron Devices*, ED-17, No. 11, pp. 975–983, November 1970.

Manuscript first received by the Institution on 28th June 1971 and in final form on 18th October 1971. (Paper No. 1429/CC 116).

STANDARD FREQUENCY TRANSMISSIONS—December 1971

(Communication from the National Physical Laboratory)

Dec 1971	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)		Dec 1971	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 60kHz		GBR 16 kHz	MSF 60 kHz	Droitwich 200 kHz	*GBR 16 kHz	†MSF 60 kHz
1	−300·2	−0·1	−0·2	562	601·8	**17**	−300·0	−0·1	−0·1	576	619·5
2	−300·2	−0·2	−0·2	564	603·7	**18**	−300·1	−0·1	−0·1	577	620·1
3	−300·0	−0·1	−0·2	564	604·9	**19**	−300·0	−0·2	−0·2	577	621·9
4	−300·2	−0·1	−0·2	566	606·2	**20**	−300·1	−0·1	−0·2	578	622·8
5	−300·1	−0·1	−0·1	567	607·1	**21**	−300·0	−0·1	−0·2	578	623·5
6	−300·0	−0·1	−0·1	567	608·0	**22**	−300·2	0	−0·1	580	623·7
7	−300·2	−0·1	−0·1	569	608·9	**23**	−300·1	−0·1	−0·1	581	623·7
8	−299·9	−0·1	−0·1	568	609·8	**24**	−299·9	0	−0·1	580	624·1
9	−300·2	−0·1	−0·1	570	611·0	**25**	−300·1	0	−0·1	581	624·5
10	−300·1	−0·1	−0·1	571	611·9	**26**	−300·2	−0·1	−0·1	583	625·2
11	−300·2	−0·1	−0·2	573	612·7	**27**	−300·1	0	−0·1	584	625·6
12	−300·0	−0·1	−0·2	573	613·2	**28**	−300·0	−0·1	−0·1	584	626·3
13	−300·1	0	−0·2	574	613·3	**29**	−300·1	0	−0·1	585	626·0
14	−300·1	−0·1	−0·2	575	613·9	**30**	−300·0	0	−0·1	585	626·4
15	−300·0	−0·1	−0·1	575	620·5	**31**	−300·1	0	−0·1	586	626·6
16	−300·1	−0·1	−0·1	576	619·7						

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}.
* 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.

Changes to the UK Standard Frequency and Time Services on 1st January 1972

In accordance with the previous notice published in *The Radio and Electronic Engineer* for November 1971 (p. 492) a time step of —0.107 760s was applied to the time standard at Rugby Radio Station, by delaying the emission of the minute pulse following 23h 59m 59s on 31st December 1971 by this amount. The step was applied to all the MSF time signal emissions and the GBR 16kHz transmissions. In addition the offset of —300 parts in 10^{10} was eliminated from the GBR signal frequency simultaneously with the time step.

The Bureau International de L'Heure, has announced that from 1st January 1972 the value for DUT 1 is —0.1s until further notice. Users of the time service will observe that the ninth second following the minute is emphasized with a double pulse, in agreement with the C.C.I.R. code for indicating the value of DUT 1.

U.D.C. 621–52

A Modular Approach to Sequential Control

ALAN PUGH,
B.Sc., Ph.D., C.Eng., M.I.E.E., F.I.E.R.E.*
and
MICHAEL E. WOODWARD,
B.Sc., Ph.D.†

Based on a paper presented at the Conference on Electronic Control of Mechanical Handling held in Nottingham from 6th to 8th July 1971.

SUMMARY

The paper gives a qualitative treatment of a novel approach to the production of sequential control systems for mechanical handling devices. A proposition that complex sequences can be broken down into combinations of four standard sequence packages is justified and methods are evolved for interconnecting these standard-sequence packages to form complete systems. Additional benefit in the form of freedom from electrical interference is gained by using the machine actuators alone as memory devices.

* *Department of Electrical and Electronic Engineering, University of Nottingham.*

† *Department of Electronic and Electrical Engineering, Loughborough University of Technology; previously of the Department of Electrical and Electronic Engineering, University of Nottingham.*

1. Introduction

Over the past four years, the combined efforts of the Department of Production Engineering and Production Management and the Department of Electrical and Electronic Engineering at the University of Nottingham have resulted in the evolution of a range of mechanical handling devices associated with a novel method of sequential control. This activity has ultimately led to the manufacture of a number of mechanical assembly machines for various customers and the eventual adoption of the system by a manufacturer specializing in the field of mechanical handling.

A brief investigation of design techniques currently used in control systems for sequential machines reveals that the availability of cheap and reliable integrated-circuit logic packages has not inhibited manufacturers from retaining and marketing discrete component systems. There are, perhaps, two reasons for this; one is based on the fact that manufacturers not only sell logic hardware, but also intuitive design techniques and rules-of-thumb for assembling their particular logic devices into a system. The designer of integrated-circuit systems is required to possess a more comprehensive knowledge of logic systems. The second is a more profound and reasonable reservation. Normal ranges of integrated circuits are fast and susceptible to malfunction in the presence of transient electrical noise of the type encountered in industrial environments. While a logic operator might have its input or output logic level disturbed only momentarily, a memory element is likely to retain its disturbed state and consequently cause continued malfunctioning of the subsequent operation of the system. Intuitive designs rely heavily on the use of memory devices of one type or another and consequently are likely to suffer from malfunctioning if the memory components are realized in integrated-circuit form.

This paper gives a brief and qualitative description of a completely new approach to the design of control systems for sequential processes. Only the combinational logic parts are realized in integrated-circuit form with the memory function of the system being provided by the mechanical actuators of the machine. This inherently removes from the control system its susceptibility to transient electrical noise due to the fact that the mechanical actuators, and hence the memory devices in the machine cannot respond to such rapid changes in signal level. The noise immunity of the system is sufficient to remove the need for noise filtering circuits from limit-switch signals and even from the integrated-circuit power supplies! In addition, the designer is no longer required to possess either a detailed or, perhaps worse, an inadequate knowledge of logic operators and their algebra. Each system is designed from its sequential properties alone, using a small number of elementary standard sequences. In addition, design times of several days using existing techniques can be reduced to just a few hours. A more detailed explanation of the philosophy is already available.[1–3]

2. The Concept of a Standard Sequence

The evolution of a number of standard sequences requires parallel thinking in both the mechanical handling applications as well as the electronic controller. It has been stated that the use of electronic memory circuits either as actual devices or as feedback loops is to be avoided in order to improve the security of operation in the presence of electrical noise. This has been shown to be possible[3] providing the sequence executes a unit-distance (cyclic) code. At this stage, it might seem that the restriction of sequences to a unit-distance nature is prohibitive, due to the fact that a great many mechanical operations are non-cyclic. However, a later section shows how it is possible to accommodate all sequences inside a unit-distance concept.

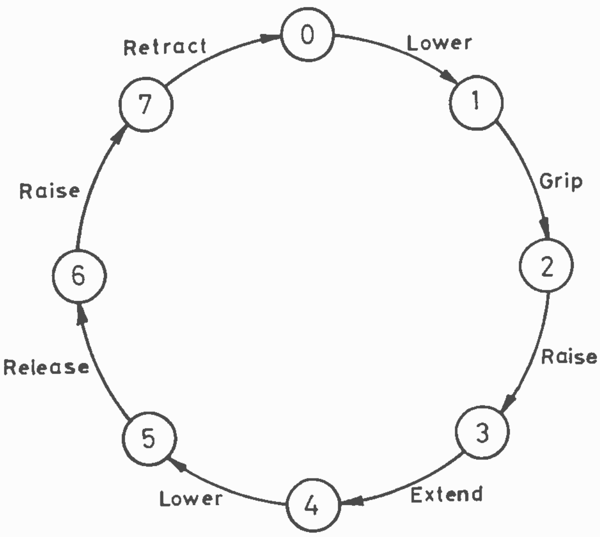


Fig. 1. Arrow diagram for an 8-step, 3-actuator placement sequence.

If an examination or survey is conducted of elementary mechanical operations, it becomes apparent that an assembly machine breaks down quite naturally into separate mechanisms containing no more than three actuators. It is accepted that 'an actuator' might be realized in practice by two or more actuators operating

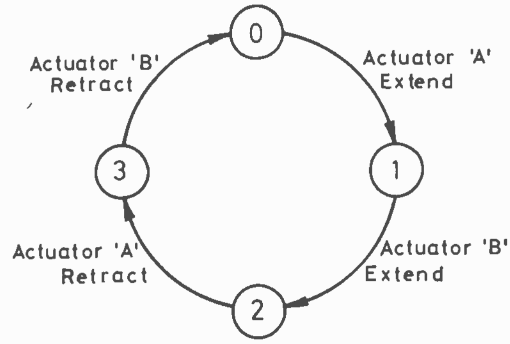
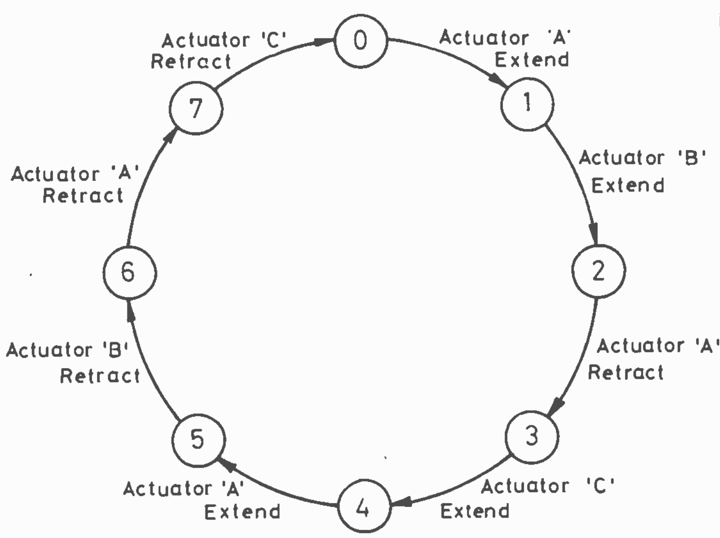


Fig. 3. Arrow diagram for a generalized 4-step, 2-actuator sequence.

in parallel. This fact in no way affects the more general concept of 'an actuator'. As an example, a sequence involving parts placement might require eight steps and use three actuators, as illustrated by the arrow diagram of Fig. 1. A fourth actuator, escaping parts into a pick-up position, could operate in parallel with the 'transfer' actuator. This is a detail of design which does not alter the fact that the sequence is made up of eight steps and three 'generalized' actuators. It is now possible to introduce one of the new terms in the philosophy. The sequence just described will be known as an '8-step, 3-actuator sequence'. Figure 2 shows the same sequence, with more general labelling, together with a table listing the actuators which are operated at each step in the sequence. It is immediately obvious that this particular sequence is cyclic and therefore conforms to the fundamental requirement of the philosophy.

An examination of other mechanical examples will reveal cyclic sequences involving three actuators, but possessing only six steps. For two actuators, the only possible choice will be of sequences involving just four steps. The generalized 2-actuator, 4-step sequence of Fig. 3 is used to demonstrate the elementary process of logic design for a standard sequence.



Position in sequence	State of actuators		
	A	B	C
0	0	0	0
1	1	0	0
2	1	1	0
3	0	1	0
4	0	1	1
5	1	1	1
6	1	0	1
7	0	0	1
0	0	0	0
ETC			

Fig. 2. The arrow diagram of Fig. 1 with generalized labelling together with a table showing the cyclic nature of the *state* of the actuators. (The '0' and '1' states indicate actuator retracted and extended respectively.)

Each actuator is associated with two limit-switches which for actuator A are labelled A_r and A_e (retracted and extended respectively). It will be assumed that the logic level 1 is used both for limit-switch operated and for the existence of a drive signal to the actuators. Table 1 gives a truth table defining the drive to each of the two actuators A and B in terms of the states of the four limit-switches. Included between each step in the sequence is the 'in-flight' condition where neither limit-switch is operated during the extension or retraction of that actuator. The four terminal steps in the sequence give rise to eight logical steps from which it is possible to derive a logic circuit to generate the sequence. This particular sequence, together with its standard-sequence controller, will form one of the building bricks of the system.

Table 1

Step in sequence	Limit switch state				Drive to actuator	
	A_r	A_e	B_r	B_e	A	B
0	1	0	1	0	1	0
	0	0	1	0	1	0
1	0	1	1	0	1	1
	0	1	0	0	1	1
2	0	1	0	1	0	1
	0	0	0	1	0	1
3	1	0	0	1	0	0
	1	0	0	0	0	0
0	1	0	1	0	ETC	

3. Enumeration of Standard Sequences

A rigorous theoretical analysis of cyclic sequences[3] reveals that there are 2 possible four-step sequences, 24 six-step sequences and 12 eight-step sequences. This means that there is a total of 38 cyclic sequences for two and three actuators. Continuing to sequences of four actuators reveals a total of 16 872 cyclic sequences! Clearly, there is an obvious practical reason for 'drawing the line' after 3-actuator standard sequences which reinforces the conclusions of the survey of elementary mechanisms.

A further examination of the 38 sequences shows that many are derivations of others merely by negating the drive to an actuator (and hence commutating its limit-switches) and/or choosing the start position at a point in the sequence other than step 0. All 38 sequences can in fact be derived from one 4-step, 2-actuator sequence, two 6-step, 3-actuator sequences and one 8-step, 3-actuator sequence, making four fundamental standard sequences in total. The sequences of Figs. 2 and 3 define the eight and four-step standard sequences respectively and the two 6-step sequences of Fig. 4 complete the range of standard-sequence modules.

A practical system can be evolved based on the existence of four separate standard-sequence controllers. However, continued logic design will produce a single logic circuit with four control (jump) terminals which will allow all four standard sequences, and hence all 38 cyclic sequences, to be generated from just one circuit card containing twelve integrated-circuit packages. The practical implications of this development are obvious, providing it is possible to couple together the standard sequences allocated to each separate mechanical mechanism to form a complete and integrated machine.

4. Provision of Interlock (Control) Facilities

The standard sequences described above have a certain degree of elegance in their own right, but are completely worthless unless provisions are made to allow the introduction of incoming and outgoing control signals. If they are to be suitable in all applications, it is essential to make available at each step or node in the sequence an output signal to indicate to associated equipment that the sequence has advanced to that step. The initiation of a drilling operation would be an example for this need.

Of even greater importance is the requirement for an arrest or hold signal which is incoming to the sequence package at each step which allows the sequence to continue only if that signal is present. These control or interlock signals are known as the 'k' terminals and the former outgoing control signals as the 'y' terminals. The k and y terminals are labelled according to their sequential position. For example, step 2 in a sequence will be associated with the control terminals k_2 and y_2 and an 8-step sequence will possess a total of 8 k terminals and 8 y terminals. Figure 5 shows the four-step sequence of Fig. 3 with its full complement of eight control terminals.

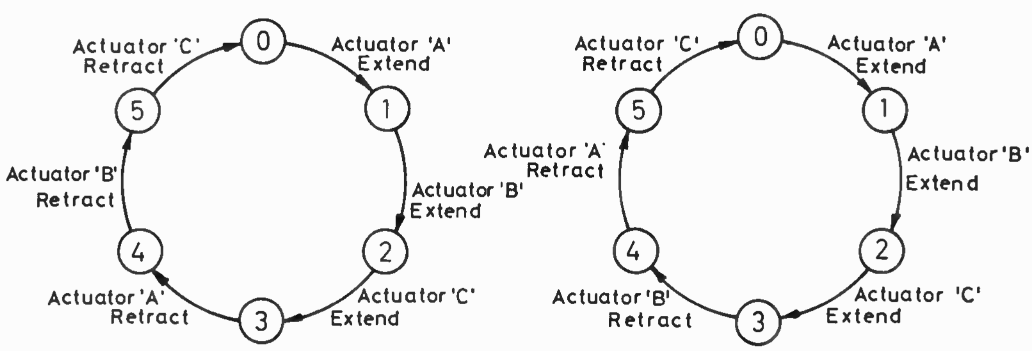


Fig. 4. Arrow diagrams for the two 6-step, 3-actuator sequences. These sequences, together with those of Figs. 2 and 3, define the four fundamental standard sequences.

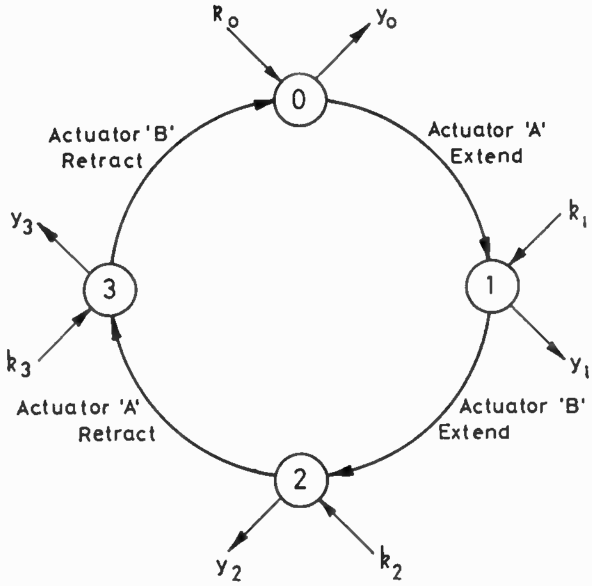


Fig. 5. Sequence of Fig. 3 showing the 'k' and 'y' control terminals.

The generation of the y signals is accomplished by simply AND-ing together the appropriate combination of limit-switch outputs. As for the k signals, it is necessary to contain these in the entries specifying the drive signals to the actuators of the mechanism. As an example, Table 2 shows the 4-step sequence of Table 1, modified to contain all four k signals together with the logical computation of the four y signals. A k signal is entered into the truth table in its true or negated form, depending on the expected change in the logic level specifying the drive to the actuator concerned. A more detailed explanation of this feature is available elsewhere.[1,2] A logic design for the 4-step sequence of Table 2 and Fig. 5, interpreted in terms of NAND elements, is illustrated in Fig. 6. This circuit is drawn from equations obtainable from Table 2. Of the 16 possible states for the four limit switches, only 8 are specified which allows some degree of freedom in the simplification process.

The existence of k and y terminals at each step in the sequence provides a very important feature when a standard sequence is associated with other equipment. A simple control such as a START button is merely connected to k_0, for example, so that the sequence will only advance from step 0 if the START button is depressed, and will always return to step 0 no matter when the START button is released. Similarly, inspection signals can be incorporated into the sequence with equal ease by utilizing the k terminals.

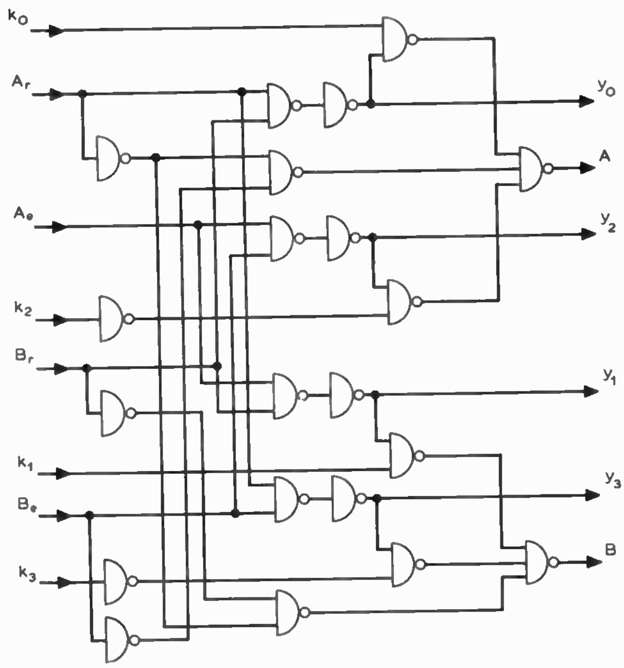


Fig. 6. Realization of the 4-step, 2-actuator standard sequence package using NAND logic elements.

5. Interconnexion of Sequences

The value of the k and y terminals becomes even more significant when two or more standard sequences are coupled together. A typical assembly or second operation machine will involve the operations of placement, fastening and deformation. In a breakdown of the mechanical operations of the machine, it becomes natural to allocate one standard sequence to the placement

Table 2

Step in sequence	Limit switch state				Drive to actuator		'y' signal
	A_r	A_e	B_r	B_e	A	B	
0	1	0	1	0	k_0	0	$y_0 = A_r B_r$
	0	0	1	0	1	0	
1	0	1	1	0	1	k_1	$y_1 = A_e B_r$
	0	1	0	0	1	1	
2	0	1	0	1	$\overline{k_2}$	1	$y_2 = A_e B_e$
	0	0	0	1	0	1	
3	1	0	0	1	0	$\overline{k_3}$	$y_3 = A_r B_e$
	1	0	0	0	0	0	

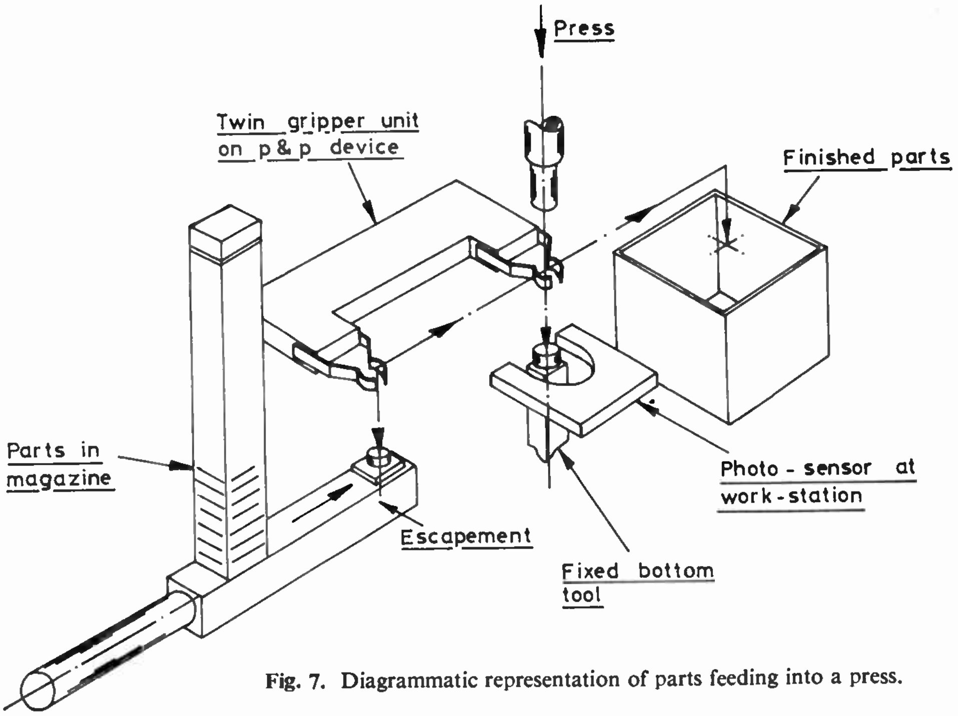


Fig. 7. Diagrammatic representation of parts feeding into a press.

mechanism and a second standard sequence to the fastening or deforming device. A simple example of such a machine is illustrated diagrammatically in Fig. 7, where a part, after being escaped into a pick-up position, is placed on a press tool for a second operation. The placement mechanism conveniently forms an 8-step sequence and the escapement actuator together with the press actuator combine to form a 4-step, 2-actuator standard sequence.

The two standard sequences are shown in Fig. 8 with labelling specific to the application. Interconnexion of the two standard sequences allows the placement and press deformation mechanisms to be completely and safely interlocked with each other. Upon the arrival of a start signal at k_0, the placement sequence advances to step 6 where it halts, because the signal y_{10} from the press sequence is not available at the terminal k_6. However, the signal y_6 enters the press sequence as k_7

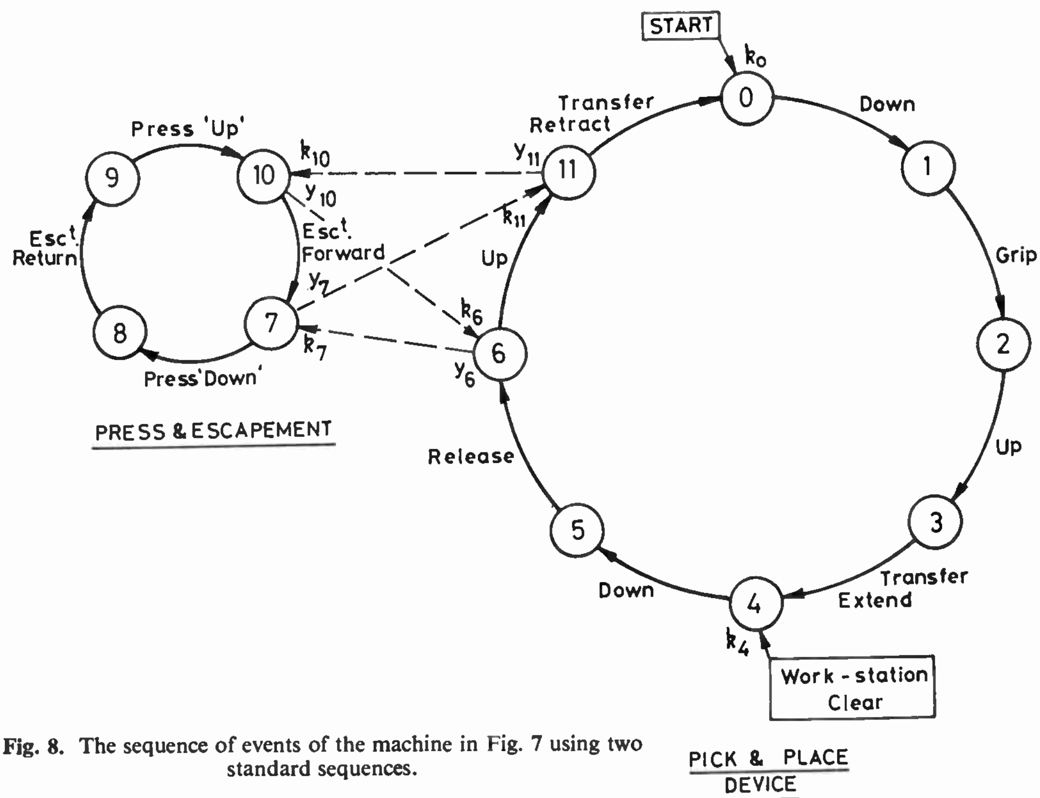


Fig. 8. The sequence of events of the machine in Fig. 7 using two standard sequences.

and allows the deformation operation to be accomplished up to step 10. The signal y_{10} now becomes available and the placement sequence advances to step 11, making y_{11} available to return the press sequence to its initial state at step 7. Finally, y_7, entering the placement sequence as k_{11}, permits the placement sequence to continue to execute another cycle.

There is no limit to the number of steps embraced by the interconnexion, and it is possible to allow both sequences to cycle independently, where they do not interfere mechanically, by computing a y signal over a section of a sequence rather than at a single step. Again, if two or more mechanisms are required to be operated in 'parallel', only a single standard-sequence controller is required. The actuators of the different mechanisms are energized from the same output signals of the sequence controller and the feedback positional information from the limit-switches are AND-ed together at corresponding actuator positions, to form a single feedback signal.

In practice, it has proved to be a relatively simple matter to break down a complex assembly or second-operation machine into elementary mechanisms in order to allocate standard sequences to each part of the machine. Indeed, the mechanical designer is often aided by this approach in understanding the sequential properties of the machine being developed.

6. Non-cyclic Sequences

It has been stated earlier that many mechanical mechanisms are non-cyclic in their operation. Referring to the example of Figs. 7 and 8, it is apparent that in the absence of the escapement actuator, the deformation sequence is reduced to a single actuator for which there is no standard sequence. In the case of a single actuator operating in isolation, it is easy to see that there is no sequential or positional difference between the times when the actuator is about to operate and when it has just completed its operation. In the absence of a memory device of some sort, the sequence is ambiguous.

This situation is accommodated in the current design philosophy by deliberately inserting an additional (auxiliary) actuator. It has been shown[3] that it is always possible to make a sequence cyclic by including one or more auxiliary actuators. For many designers of mechanical handling machines, it is not unreasonable for them to include an actuator which is doing no more than operating its two limit-switches. Again, an improvement in mechanical design can be obtained with this additional actuator by the facility of being able to control the time of operation of the auxiliary actuator. In the case of the press, time would have to be allowed for the part to be properly and regularly deformed. In some applications, the auxiliary actuator can be used to perform a 'timed' task such as 'firing' a resistance welder.

The press sequence of Fig. 8 would, therefore, be made cyclic by including an auxiliary actuator in place of the escapement, and hence preserving the sequential property of the cycle of operations. Where it is preferred, an electronic actuator simulator is available to use in place of a physical auxiliary actuator. This is merely a delay circuit providing simulated limit-switch signals at some variable time after the drive signal is established and removed. A third alternative for an auxiliary actuator is an electromagnetic relay with one changeover contact. This combines an actuator with two limit-switches and hence satisfies the requirements for use in the system.

The concept of auxiliary actuators to convert non-cyclic sequences to cyclic and hence standard sequences completes the philosophy of the system.

7. Associated Hardware

In terms of physical hardware, the standard-sequence controllers form only a part of a complete control

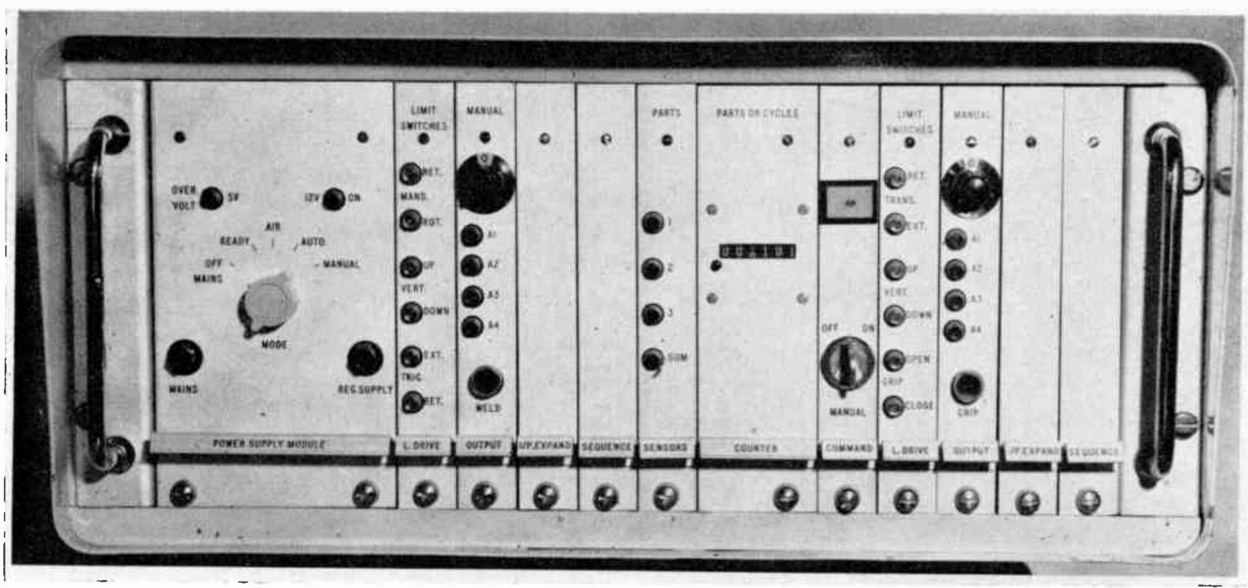


Fig. 9. A typical complete control system for an assembly machine.

system. A typical installation is illustrated in Fig. 9 which contains a power supply, solenoid drivers, limit-switch indicators, photosensor amplifiers, parts counting and air interlock modules. All these have been designed to be compatible with the standard-sequence philosophy. For example, the solenoid driver module contains all the controls for manual operation which are operational only when the machine is switched to manual. A removable key is included for this facility so that its operation can be restricted to a supervisory staff grade.

(By courtesy of Brush Electrical Engineering Co. Ltd., Loughborough)

Fig. 10. A 2-part assembly machine showing general arrangement of mechanisms, sequential control system and operator's controls.

In addition to the controls included on the modules themselves, it is normal to repeat the essential start/stop controls near the operating position. The operator's start control also embodies single or half-cycle operation by directing its signal to one or more k terminals in the operating sequence. To run the machine continuously, it is necessary to maintain pressure on the start button over the entire first cycle. The 'auto-run' relay will then latch near the completion of the first cycle. This facility gives to the operator extensive control over the machine with the minimum number of control buttons.

A typical assembly machine together with its control system is illustrated in Fig. 10. The actuators of this machine are both pneumatically and hydraulically powered and the placement device is one of the range of mechanisms designed to integrate with the philosophy of standard sequences.

8. Operating Experience

Most of the machines installed in industry have been in production for a year or more with a high degree of success. The control systems have proved extremely reliable with no tendency to malfunction in the presence of electrical interference. In particular, the installation of Fig. 10 is associated with a heavy current ($\sim$ 1000 A) welding machine, without any operating difficulties whatsoever in the control system

There is also considerable evidence that the costs of both the control system hardware as well as the design time is much reduced, and savings in excess of 50% in total control system cost have been achieved over conventional techniques. It is the authors' hope that these techniques will be accepted in all areas of mechanical handling where sequential control is all that is required.

9. Acknowledgments

The authors wish to thank the University of Nottingham, in particular Professor W. B. Heginbotham, Head of the Department of Production Engineering and Production Management, for providing facilities to carry out this work. The entire mechanical design of the handling devices is due to a team headed by Dr. J. R. Ashley, and much of the detail design of the control systems is the responsibility of a team under the guidance of Mr. J. E. Eziashi.

Hawker Siddeley Dynamics Limited, Hatfield, Herts., England, are now holders of exclusive rights to market control systems based on this design philosophy[4] and currently manufacture the 'Minitran' range of placement devices using such a control system.

10. References

1. Ashley, J. R. and Pugh, A., 'Logical design of control systems for sequential mechanisms', *Internatl J. Prod. Res.*, **6**, No. 4, 1968.
2. Ashley, J. R., Pugh, A. and Woodward, M. E., 'Synthesis of complex sequential control systems from standard-sequence packages', *Internatl J. Prod. Res.*, **9**, No. 3, 1971.
3. Woodward, M. E., 'Theory and design of switching circuits for the sequential control of machine operations', Ph.D. thesis, University of Nottingham, December 1970.
4. 'Sequential mechanisms', British Patent Application No. 00800/70, Hawker Siddeley Dynamics Limited.

Manuscript first received by the Institution on 29th March 1971 and in final form on 4th November 1971. (*Paper No. 1430/CC 58.*)

Contributors to this issue

Mr. Leo Ojala received the Diploma Engineer and Licentiate of Technology degrees at Helsinki University of Technology in 1960 and 1965. From 1959 to 1962 he was a nuclear electronics engineer in the Finnish Institute of Radiation Physics, and he then joined the Finnish Atomic Energy Commission as a research assistant, working mainly on ultra-fast logic circuits until 1968. He then joined the Department of Electrical Engineering of Helsinki University of Technology and was responsible for initiating the research work in special-purpose computers and digital processing of signals. In 1971 he became an Acting Professor of Digital Techniques of the University.

Mr. Pentti Lappalainen obtained the Diploma Engineer degree in electronic engineering at Helsinki University of Technology in 1969. Since then he has worked in the Applied Electronics Laboratory of the University as a member of a team carrying out research into new solutions of television test equipments, special applications of television techniques and picture processing by computer.

Mr. E. Howbrook started his technical training as an apprentice at Metropolitan-Vickers and after completing his apprenticeship entered the electronic control department as a technician, later becoming an engineer. He left the company to enter the guided weapons industry, first as a senior engineer at English Electric, Luton, then at the Royal Radar Establishment, Malvern, and finally, from 1954 until 1964, as a section leader at Hawker Siddeley Dynamics, Woodford. After the completion of the company's contract he joined the Department of Electrical Engineering of the University of Salford.

Mr. K. J. Glover obtained a first class honours degree in electrical engineering at King's College, University of Durham, in 1962. He then worked in the English Electric Company's Computing Division before joining Bradford University as a lecturer in 1963. He obtained an M.Sc. from Newcastle University in 1967.

(Biographies of Professor D. P. Howson and Dr. J. G. Gardiner were published in September 1969 and May 1969 respectively.)

Mr. F. C. Evans (Member 1963) graduated with a B.Sc. honours degree in physics from University College London in 1953, having previously completed two years National Service as a radar fitter in the Royal Air Force. He spent two years in industry with the Engine Division of Bristol Aeroplane Company and then moved into teaching with part-time research, first at Birmingham College of Advanced Technology (1955–59) as an assistant lecturer and then at Kingston College of Technology (1959–66) as lecturer and later senior lecturer. Mr. Evans joined the University of St. Andrews in 1966 as lecturer in physics and became senior lecturer in 1971. His main research interest is in practical implementation of pattern recognition systems.

Dr. A. Pugh (Fellow 1970; Member 1965) obtained a first class honours degree in electrical engineering at the University College of South Wales and Monmouthshire in 1957. After graduation, he joined the B.B.C. as a graduate trainee in their specialists departments where he stayed until 1959, when he moved to Rolls-Royce and Associates Ltd., as a nuclear instrumentation engineer on the British nuclear submarine programme. He took up his present position as a member of the academic staff in the Department of Electrical and Electronic Engineering at the University of Nottingham in 1961. In 1968 he was awarded a doctorate for work on non-binary codes applied to digital control systems. Dr. Pugh is currently a supervisory member of a team working in the field of mechanical handling devices with visual capability. He served on the organizing committee for the I.E.R.E. Conference on Electronic Control of Mechanical Handling and is currently a member of the Components and Circuits Group Committee. A previous paper on a positioning system using three valued codes gained him the J. Langham Thompson Premium for 1968.

Dr. M. E. Woodward was employed by British Rail Signal and Telecommunications Department from 1960 to 1964, during which time he attended Derby and District College of Technology, obtaining a H.N.C. in electrical engineering. He graduated from the University of Nottingham with a first class honours degree in electronic engineering in 1967, and was awarded a doctorate by the University in 1970 for research into sequential control systems. He at present holds the post of Baker-Perkins lecturer in industrial instrumentation in the Department of Electronic and Electrical Engineering, Loughborough University of Technology.