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. 11 NOVEMBER 1972

THE RADIO AND ELECTRONIC ENGINEER

The Journal of the Institution of Electronic and Radio Engineers

Half a Century of Engineering Achievement

YEARS ago—on 14th November 1922—the then British Broadcasting Company inaugurated its regular service by a broadcast over the London Station 2LO from studios in Marconi House in the Strand. During the intervening 50 years —which only slightly overlap the life span of this Institution, founded in 1925— it has achieved a position among the world's broadcasting authorities which can justly be claimed to be second to none. This position of eminence applies both to its professionalism in programmes and, especially, to its engineering expertise.

The broadcast recollections of fifty years of programmes of all kinds have been supplemented by exhibitions in London: that at the Langham, opposite Broadcasting House, is orientated primarily towards the listener (and viewer) who can experience recordings of some of the notable broadcasts of each decade associated with contemporary room settings as backgrounds to typical receivers. The engineering story of British broadcasting is told half a mile away, in the complementary exhibition mounted jointly by Mullard Ltd. and the BBC and linked, to the fascination of younger visitors, by two-way colour television circuits.

This very competently staged technical exhibition at Mullard House has both historical and modern items, the former including a reconstruction of the original 2LO studio, and various microphones, amplifiers and disk, wire and tape recorders. Modern techniques include colour television equipment, Post Office exhibits associated with programme distribution and quadraphonic demonstrations. Another section of the exhibition shows such 'futuristic' developments as broadcasting from satellites, low-light-level television, holographic displays and waveguide and fibre optics transmission.

Another exhibition which lays special emphasis on the earliest days of broadcasting has been arranged by the Institution of Electrical Engineers. Many readers will recall that between 1923 and 1932 the BBC's studios and offices were in part of the IEE's building, using the address 'Savoy Hill'. For the first half of this period the (first) Chief Engineer of the BBC was P. P. Eckersley, a prominent member of the IEE and of this Institution.

This anniversary year has been marked by the publication of several books, some putting forward personal views of the way in which broadcasting has grown from a hobby activity for the listener to a major factor in popular entertainment (and enlightenment!), while others deal with the technicalities of broadcasting. Pride of place among the latter must be given to the immensely authoritative and intensely interesting 'BBC Engineering 1922–1972' written by Edward Pawley, who recently retired after 40 years with the Corporation, latterly as Chief Engineer, External Relations. The Science Museum has also published a booklet 'Broadcasting in Britain 1922–1972', by Keith Geddes, which gives a brief and well illustrated account of the major engineering developments.

The early and glorious days of broadcasting have been admirably captured in a film made by Mullard Ltd., which covers the years up to the Coronation in 1953. Its title 'Cough and you'll deafen thousands' is a quotation from the 'instructions' placed on the microphone! With its reminiscences by many of those who took part in laying the technical foundations of British broadcasting, and who later held senior posts, it is a truly fascinating and amusing 52 minutes.

Seldom can a public anniversary have been celebrated so extensively and it is difficult to find original comment. The IERE has always enjoyed excellent relations with the BBC, many of whose engineers are members, a goodly proportion serving on Council or its Committees, and numerous papers have been contributed to the *Journal* or read at meetings over the years. It is, therefore, with real sincerity that on behalf of all its members the Institution tenders congratulations to the British Broadcasting Corporation.

* The BBC–Mullard Exhibition is open until 21st December; the IEE Exhibition until 30th November.

Contributors to this issue*

Mr. R. G. Bennetts received a B.Sc. in aeronautical engineering in 1966 from the Council for National Academic Awards and an M.Sc. in electronics in 1970 from Southampton University. From 1966 to 1969 he was a research engineer with the Plessey Company at Roke Manor, where he was responsible for designing and testing part of a data processing system. He has just completed a period of research in the Department of Electronics at Southampton University and during this period he was mainly concerned with techniques for generating testing sequences for logic circuits. Other research activities have been concerned with the use of switching theory in connection with logic design, and he has published papers in both these areas. He is now a Lecturer in the Department of Electronics at Southampton University.

Professor D. W. Lewin (Member 1960, Graduate 1957) was appointed to the chair of electronics at Brunel University in January of this year. During the previous five years he was a Lecturer and subsequently a Senior Lecturer in the Department of Electronics at the University of Southampton; from 1962-67 he lectured in computer engineering at Brunel University. Professor Lewin, who is the author of several books and numerous papers, has been a member of the Papers Committee since 1971 and the Computer Group Committee since 1970 and he was for several years a member of the Southern Section Committee. He has represented the IERE on the Organizing Committees of several joint conferences.

Mr. J. L. Washington is a Senior Programmer in the Department of Electronics at Southampton University. He obtained a B.A. in mathematics at St. John's College, Cambridge in 1965, and in 1967 an M.Sc. in electronics at Southampton University. Since then he has been engaged in research on speech analysis and recognition, and, since 1970, switching theory and computer-aided logic design.

Mr. R. W. J. Barker and **Mr. B. L. Hart** (Member 1961, Graduate 1955) are Senior Lecturers at Portsmouth and North East London Polytechnics respectively. (See *Journal* for March 1972.)

Mr. R. J. Wescott is Head of the Data Transmission Section in the Post Office Research Department, Dollis Hill, which he joined in 1954, after having worked for nine years in the Exeter Telephone Area. Between 1954–1958 he was concerned with the development of steerable aerial systems for h.f. radio and over the next three years with low-loss circular waveguide systems. Mr. Westcott worked on various aspects of satellite communication systems between 1962 and 1969 including low-threshold demodulators, low-loss waveguide systems, analysis of experiments for satellites *Telstar*, *Relay* and *Early Bird*, and general system studies for the *Intelsat* satellites such as multi-carrier transmission via travelling-wave tubes and digital transmission systems etc. Since 1969 he has been concerned with the development of data transmission systems.

Dr. V. A. Cherdyntsev obtained a B.Sc. degree in electrical engineering from Sverdlorki Polytechnic Institut, USSR, in 1959 and a Ph.D. degree (Candidate of Science) from Moscow Energetic Institute in 1965. He joined Minsk Radiotechnical Institute in 1968 where at present he is the 'head of the chair' of technology and radio engineering. His main research interests include problems of digital communication, processing and filtering and he participated in the work being carried out in these subjects at City University while on study leave.

Dr. A. R. Memon graduated in 1964 with honours in physics from Sind University, Pakistan. He obtained an M.Sc. in electronic circuit and system engineering from Bath University of Technology in 1967. Following research work at City University he has been awarded a Ph.D. for his thesis on 'Synthesis of linear phase recursive digital filters'. In addition to digital filters his research interests include computer-controlled telecommunication systems.

Mr. M. A. Khan (Graduate 1968) received a B.Sc. (Hons.) degree in electrical engineering from the Polytechnic of Central London in 1969; he then joined the Post Office Research Station, Dollis Hill, where he was associated with the digital switching group until 1970. He is at present a research student with the Department of Electrical and Electronic Engineering at the City University on special study leave from the Post Office. His research is mainly concerned with the design and analysis of digital filters.

* See also pages 488, 495 and S.186.

U.D.C. 621.373.544: 621.397.332

Designing a Television Line Flywheel Generator Using a Phase-locked Loop Integrated Circuit

P. POMEROY (Graduate)*

SUMMARY

The adaptation of phase-locked loop theory to meet the design requirements of a television line flywheel generator is discussed. The operation of a loop driven with narrow line sync. pulses is described with regard to both 'in sync.' and 'pull-in' performance. A design example is given to illustrate the design procedure.

* *Natal College for Advanced Technical Education, Durban, South Africa.*

List of Symbols

ω_0	system input angular frequency, rad/s
ω_n	loop natural angular frequency, rad/s
$\Delta\omega_p = 2\pi\Delta f_p$	pull-in difference angular frequency, rad/s
ω_p	pull-in angular frequency, rad|s
$\Delta\omega$	angular frequency step, rad/s
θ_i	input phase, rad
θ_O	output phase, rad
θ_l	in-sync. phase error, rad
t_l	timing error corresponding to θ_l, s
θ_u	unlocked phase error, rad
t_s	relock instant, s
θ_{us}	phase error at relock, rad
t_{us}	timing error corresponding to θ_{us}, s
$\Delta\theta$	phase step, rad
$\theta_p(t)$	transient phase error due to $\Delta\theta$, rad
$\theta_f(t)$	transient phase error due to $\Delta\omega$, rad
$\theta(t)$	total transient phase error, rad
v_D, V_D, V_{DM}	phase detector output voltage, V
K_D	phase detector sensitivity, V/rad
K_O	v.c.o. sensitivity, rad/s/V
ζ	damping factor
$H(s)$	phase transfer function of loop
B_L	loop noise bandwidth, Hz
$F(s)$	transfer function of loop filter
T_S	sync. pulse width, s
T_L	line period, s
$d = T_S/T_L$	sync. pulse duty cycle

1. Introduction

The principle of flywheel line synchronization in television is well known and its ability to preserve synchronization through noise and the field sync. period is exploited in modern equipment. The flywheel principle is in fact an application of a phase-locked loop (p.l.l.) although, to the author's knowledge, it has not been viewed in this way for design purposes.

The aim of this paper is to present a design procedure for the flywheel circuit, adapting classical p.l.l. theory to meet the requirements of a loop driven by narrow sync. pulses. The fundamental principles of phaselock are presented briefly, the reader being referred to the literature for a detailed account.

A sample design, using a commercially available p.l.l. i.c. (National Semiconductor Corporation LM565) illustrates the relationships between the design criteria presented throughout the paper.

2. 'In Sync.' Performance

2.1. *The Phase-locked Loop System*

The p.l.l. has the ability to provide a signal at its output tracking the phase of a signal at the input. It is an example of a feedback control system, consisting of the three basic components shown in Fig. 1.

The voltage-controlled oscillator (v.c.o.) is of the RC-type,[1,2] the centre or free-running frequency being set by external components. The output is a t.t.l. compatible square wave which can be easily shaped for line deflexion drive or used directly in instrumentation applications, e.g. to drive a divide-by-625 divider chain. The phase detector provides the v.c.o. with a d.c. control voltage proportional to the phase difference between the system output and input. Thus, should there be any attempt to change phase on the part of either the input or the output, the d.c. control voltage serves to keep the v.c.o. tracking the input in both phase and, therefore, frequency.

The phase error voltage (v_D in Fig. 1) is filtered by the low-pass filter. The filter also determines the dynamic performance of the loop; i.e. the response to input phase or frequency changes. The filter considered in this paper is of the passive, first-order lead-lag type shown in Fig. 2.

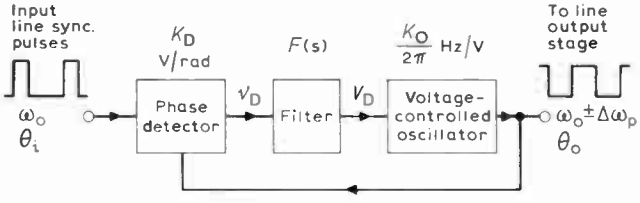


Fig. 1. The phase-locked loop system.

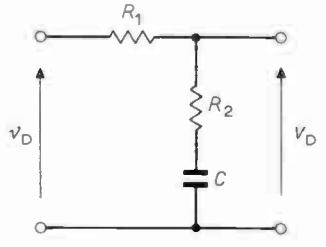


Fig. 2. The loop filter.

Two filter time-constants are defined:[1-4]

$$T_1 = R_1 C \qquad \text{.....(1)}$$

$$T_2 = R_2 C \qquad \text{.....(2)}$$

It is easily shown[2-4] that the v.c.o. output phase θ_O resulting from a change in system input phase θ_i is

$$\frac{\theta_O(s)}{\theta_i(s)} = H(s) = \frac{K_O K_D F(s)}{s + K_O K_D F(s)}, \qquad \text{.....(3)}$$

where $F(s)$ is the filter transfer function and $K_O K_D$ is the loop gain (see Fig. 1).

2.2. *The Phase Detector*

The phase detector in Fig. 1 is effectively a switch[2] that inverts the system input on alternate half cycles of the v.c.o. output. The operation is illustrated in Fig. 3 with the loop locked and the phase difference being such as to give zero mean d.c. control voltage to the v.c.o.

The phase detector inverts the input waveform on positive half cycles of the v.c.o. signal. Figure 3 shows that when the input pulse centres coincide with the

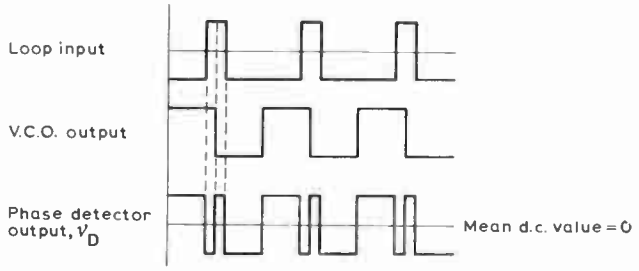


Fig. 3. Phase detector operation.

negative-going edges of the v.c.o. waveform, the phase detector interprets this as a phase difference of zero. It is therefore convenient to define an 'in-sync.' phase error θ_l, expressed as a timing error t_l, giving the time between the pulse centres and the negative-going edges of the switching waveform.

$$t_l = \frac{\theta_l}{2\pi} \times T_L. \qquad \text{.....(4)}$$

The form of the phase detector output is shown in Fig. 4 with the v.c.o. leading the input pulses in phase, i.e. the v.c.o. attempting to 'free-run' above the input frequency.

For

$$0 < t_l < T_S/2$$

the mean d.c. value of the detector output is easily found by considering areas under the v_D waveform to be:

$$V_D = 4V_{DM} t_l / T_L.$$

Using the duty cycle,

$$d = T_S/T_L,$$

this can be rewritten:

$$V_D = 2dV_{DM} \frac{t_l}{T_S/2}. \qquad \text{.....(5)}$$

When t_l exceeds $T_S/2$, V_D becomes constant until t_l has such a value that the positive-going edges of the v.c.o. output coincide with the trailing edges of the sync. pulses. This occurs when

$$t_l = \frac{T_L}{2} - \frac{T_S}{2},$$

and can be rewritten

$$t_l = \frac{T_L}{2}(1 - d).$$

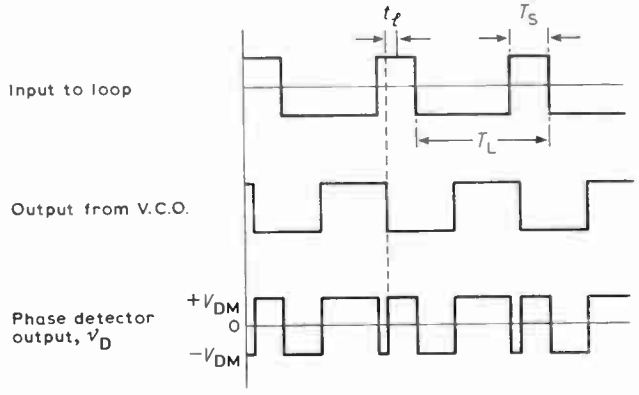


Fig. 4. Phase detector operation with a timing error, t_l

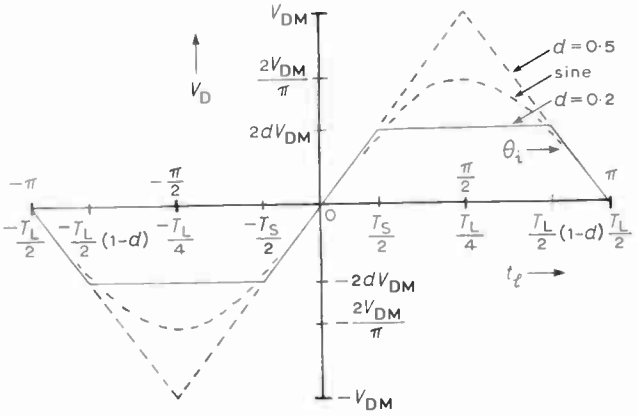


Fig. 5. Phase detector sensitivity.

Thus the range for which V_D is constant is

$$T_S/2 < t_i < T_L/2(1-d),$$

giving V_D a value

$$V_D = 2dV_{DM}. \qquad \text{.....(6)}$$

A similar result is obtained for lagging phase. The form of the phase detector control characteristic is shown in Fig. 5 for both a pulse and a sinusoidal loop input. (v_D of Fig. 4 becomes a full-wave rectified sine-wave[2] at $\theta_i = \pi/2$.)

The slope of the curve of Fig. 5 is the phase detector sensitivity K_D in volts per radian. It should be noted that at $\theta_i = 0$, K_D has the same value regardless of the input waveform, and at $\theta_i = \pi$ the sensitivity is negative and will tend to drive the loop out of lock.

2.3. *Static Phase Error*

When a phase error exists with the loop locked, a control voltage is fed to the v.c.o. and implies therefore that the v.c.o. is being held away from its free-running frequency. It has been shown elsewhere[2,3] that an initial detuning error $\Delta\omega$ of the v.c.o. will still preserve lock, with a 'static' phase error of $\Delta\omega/K_O K_D$ radians. This phase error corresponds to a change of the line sweep triggering instant and leads therefore to a horizontal shift of the picture. It is shown in Section 5 that the maximum tuning error is the pull-in frequency $\Delta\omega_p$, giving the limits of the v.c.o. free-running frequency as

$$\omega_0 \pm \Delta\omega_p.$$

Also here, the v.c.o. sensitivity K_O is proportional to the free-running frequency so that, in terms of the design value of loop gain $K_O K_D$, we get an effective loop gain of

$$\frac{\omega_0 \pm \Delta\omega_p}{\omega_0} K_O K_D.$$

This leads to a static phase error of

$$\theta_i = \frac{\pm\Delta\omega_p}{(1 \pm \Delta\omega_p/\omega_0)K_O K_D}, \qquad \text{.....(7)}$$

showing that when the v.c.o. is tuned to $\omega_0 + \Delta\omega_p$, the picture shift (to the right) is less than for a tuning error of $-\Delta\omega_p$.

3. Noise Performance

3.1. *Loop Response to Sinusoidal Phase Modulation*

The p.l.l. can preserve lock with the input changing in phase, the nature of the output phase depending on the form of the input phase change, e.g. sinusoidal, random, step, etc. In general for any input phase change the loop transfer function, equation (3), with the filter of Fig. 2 becomes[3,4]:

$$H(s) = \frac{\theta_O(s)}{\theta_i(s)} = \frac{s\omega_n(2\zeta - \omega_n/K_O K_D) + \omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}. \qquad \text{.....(8)}$$

Equation (8) shows the output phase for any form of input phase change.

Using servomechanism terminology, ω_n is the natural angular frequency and ζ the damping factor.

$$\omega_n = \sqrt{\frac{K_O K_D}{T_1 + T_2}} \qquad \text{.....(9)}$$

$$\zeta = \tfrac{1}{2}\sqrt{\frac{K_O K_D}{T_1 + T_2}}\left(T_2 + \frac{1}{K_O K_D}\right) \qquad \text{.....(10)}$$

Now, for θ_i a sinusoid, i.e. the system input being a sinusoidally phase modulated wave, equation (8) gives $H(j\omega)$ and shows how the output phase amplitude varies with the modulating frequency. This is shown in Fig. 6.

Referring to Fig. 6 it can be seen that the loop performs a low-pass filtering operation on phase inputs, i.e. the output phase tracks the input phase exactly when modulated slowly but cannot follow the input phase when the modulation frequency is high. It is of interest to note that, for low damping, the output phase exceeds the input phase when the modulating frequency is near ω_n.

3.2. *Loop Response to Noise*

When the input phase θ_i varies at random (due to a noise voltage effectively phase modulating the input signal), the r.m.s. value of the output phase jitter depends on the area under the curve of Fig. 6. A noise bandwidth B_L which can be viewed as

$$\frac{\text{mean square output phase jitter,}}{\text{mean square input phase jitter}}$$

has been defined[2,3]:

$$B_L = \int_0^\infty |H(j\omega)|^2 \, df = \frac{\omega_n}{2}\left(\zeta + \frac{1}{4\zeta}\right). \text{Hz} \qquad \text{.....(11)}$$

Note that an instantaneous positive input phase change corresponds to an instantaneous increase in the system

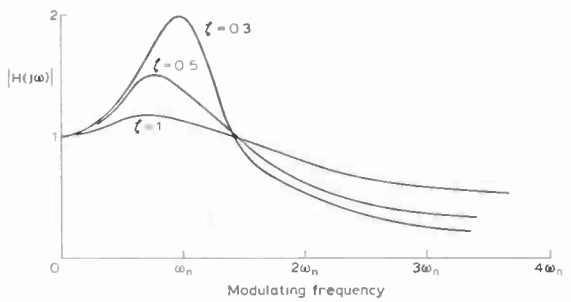


Fig. 6. Loop frequency response.

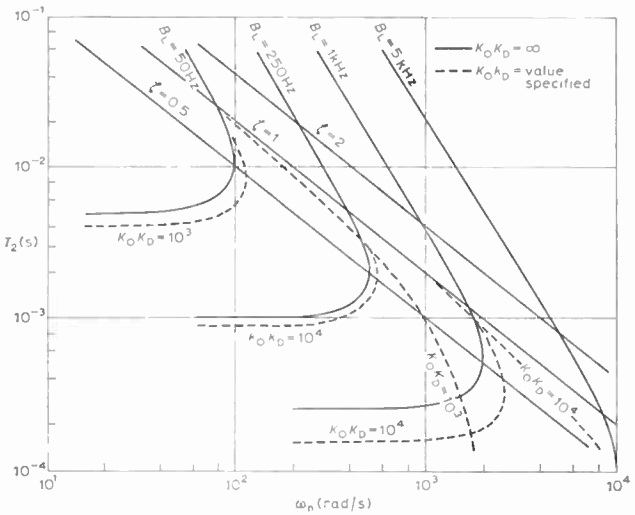


Fig. 7. Design interrelationships.

input frequency. The loop responds equally well to positive and negative input phase changes. For this reason B_L defined above has been called[5,6] the noise semi-bandwidth F_{NN}, the true noise bandwidth being $2F_{NN}$.

3.3. *Minimum Noise Bandwidth*

Equation (11) shows that noise bandwidth depends on ω_n and ζ, and to minimize output phase jitter it is desirable to design for minimum noise bandwidth. It can be shown that B_L is a minimum when $\zeta = 0{\cdot}5$ in which case, $B_L = \frac{1}{2}\omega_n$. (Note that the dimensions of ω_n are radians per second while B_L is in Hz.) Thus, having chosen ω_n to satisfy other criteria, noise bandwidth cannot be less than $\frac{1}{2}\omega_n$.

It is useful to examine how B_L, ζ, ω_n and T_2 are related. This is given in Fig. 7, together with the effect of finite loop gain, $K_O K_D$.

It would seem from the above that a damping factor of 0·5 is desirable to minimize output phase jitter. It will be seen later however that it may be necessary to design for $\zeta > 0{\cdot}5$ to achieve satisfactory relocking after field sync.

3.4. *The Raster with a Noisy Loop Input*

Phase jitter at the output of the loop causes early or later initiation of a line sweep. It is difficult to define a satisfactory figure for output phase jitter since the subjective picture quality resulting from erroneous line timing will depend on the nature of the noise and the loop constants. Consider for example a loop with a damping factor of 0·3, and a natural frequency $\omega_n/2\pi =$ 250 Hz. For white noise at the input, Fig. 6 shows that the output phase jitter will be at its worst at the natural frequency, although the line generator will be continually triggering erroneously. The largest horizontal triggering errors will therefore occur twice every 1/250th of a second, or 1/10th of picture height apart as shown in Fig. 8.

Fig. 8. The subjective effect of noise entering the loop.

3.5. *Noise Performance with Pulse Input*

In the discussion above it is assumed that the system input is phase modulated by a noise voltage. In the case of a sinusoidal signal in the presence of noise, it is easy to determine the effective phase modulation by the noise.[3,5,6] For the separated line sync. pulses accompanied by noise, the effective phase modulation depends on the rise and fall times of the pulses and the amplitude of the 'window' which the p.l.l. accepts at the input. Figure 9 shows the effect of limiting on the noise entering the loop.

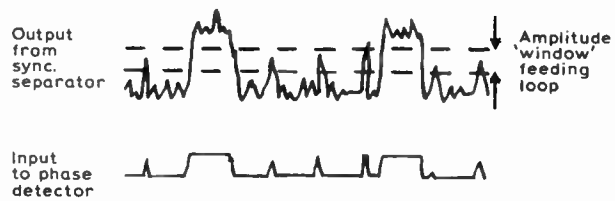


Fig. 9. Noise limiting at the loop input.

The spurious noise pulses crossing the 'window' do affect the v.c.o. output phase, but due to the integral action of the filter smoothing over the pulses, the output jitter is far less than for the same signal to noise ratio with a sinusoidal input signal.

Because of the difficulty in analysing the situation above, noise performance does not enter into the design criteria, but is assessed experimentally as a final design check. This is contrary to normal p.l.l. practice, but in this application we are less concerned with the noise-reducing capabilities of the p.l.l. than with its other properties. Here we are providing a jitter-free signal from a relatively 'clean' signal to start with, the input signal/noise ratio being 20 dB or better from considerations of the visibility of the brightness variations caused by noise on the picture signal. In other p.l.l. applications the prime consideration is often one of reconstructing a signal deeply embedded in noise.

4. Loop Behaviour during Field Blanking

4.1. *Frequency Memory*

When a broad field sync. pulse is presented to the loop, the v.c.o. 'flywheels' through this period and relocks again with the reappearance of line frequency pulses. The loop filter 'remembers' the last frequency present before the input altered and the v.c.o. continues to provide an output to the line driver. (Although the discussion is based on a single broad pulse for field

sync., analysis shows that the phase detector provides the same mean d.c. output for double line frequency equalizing and sync. pulses as for a single broad pulse.)

In the absence of a locking signal at the system input, the v.c.o. frequency drifts slowly towards its free-running value, the information being stored as a charge in C of Fig. 2. Immediately the phase detector fails to provide a control voltage to the filter, the d.c. signal to the v.c.o. experiences a step of $R_2/(R_1+R_2)\ V_D$, followed by an exponential decay with a time-constant (T_1+T_2). The v.c.o. frequency then moves exponentially towards its 'target' value. It is shown in Section 5 that the maximum possible v.c.o. mistuning is the pull-in frequency $\Delta\omega_p$. Thus the final frequency will be $\omega_0 \pm \Delta\omega_p$. The v.c.o. behaviour is shown in Fig. 10(a), for $\Delta\omega_p$ positive.

The instantaneous angular frequency ω_i can be shown to be:

$$\omega_i = \omega_0 + \Delta\omega_p \times \times \left\{ \frac{T_2}{T_1+T_2} + \frac{T_1}{T_1+T_2}\left[1-\exp\left(\frac{-t}{T_1+T_2}\right)\right]\right\}. \quad (12)$$

The instantaneous phase error θ_u, noting that the phase error has an initial value θ_l, is:

$$\theta_u = \theta_l + \int \omega_i \mathrm{d}t - \omega_0 = \theta_l + \Delta\omega_p T_1 \times \times \left\{ \frac{t}{T_1+T_2}\left[1+\frac{T_2}{T_1}\right] - 1 + \exp\left(\frac{-t}{T_1+T_2}\right)\right\}. \quad (13)$$

This is shown in Fig. 10(b).

For most loops,

$$T_1 \gg T_2$$

and in this application line drive reappears at time t_s such that

$$t_s \ll T_1+T_2.$$

Equation (12) reduces to

$$\omega_i \simeq \omega_0 + \Delta\omega_p \left(\frac{t+T_2}{T_1}\right) \quad \text{.....(12a)}$$

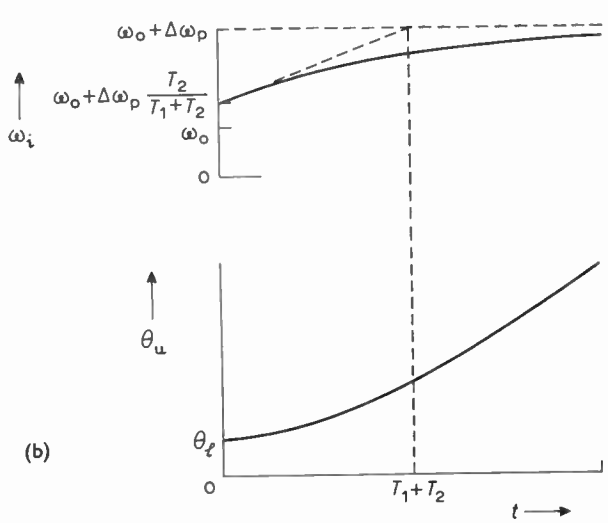


Fig. 10. Frequency and phase drift during field sync.

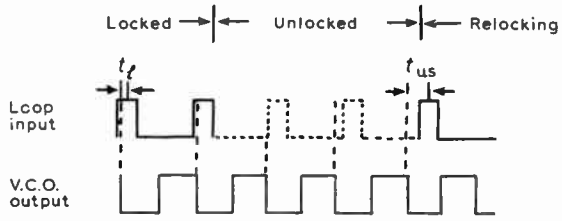


Fig. 11. V.C.O. frequency drift during field sync.

which gives

$$\theta_u \simeq \theta_l + \Delta\omega_p \frac{t}{T_1}\left(\frac{t}{2}+T_2\right). \quad \text{.....(13a)}$$

We have now established that the v.c.o. gains both a frequency and phase error during the field sync. interval.

4.2. *Relocking after Field Sync.*

4.2.1. The initial conditions at relock

When line drive reappears after field sync. the loop responds to the v.c.o.'s error as if a step in both frequency and phase is applied to the system at the relocking instant.

Figure 11 shows how the v.c.o. frequency gains on the input during a loss of input drive, assuming the mistuning to be $+\Delta\omega_p$.

In the locked condition, the timing error t_l corresponds to the static phase error θ_l discussed in Section 2.3. The timing error t_{us} corresponds to the total phase error which the v.c.o. gains over the input before the input is restored. If this total error t_{us} is more than $T_S/2$ (as indicated) Fig. 5 shows that the phase detector will be 'saturated'. This implies a reduction in the effective loop gain. Thus to achieve relocking with maximum loop gain, it is necessary that

$$t_{us} < T_S/2$$

Expressing this as a phase angle θ_{us} and using

$$d = T_S/T_L$$

we get as an initial requirement for successful relocking

$$\theta_{us} < d\pi \quad \text{.....(14)}$$

Consider now that the v.c.o. starts the frequency drift with an initial phase error of

$$\theta_l = \frac{\Delta\omega_p}{(1+\Delta\omega_p/\omega_0)K_O K_D}. \quad \text{.....(7)}$$

Figures 10 and 11 and equation (13a) show that the v.c.o. has gained a phase error $\Delta\theta$ of

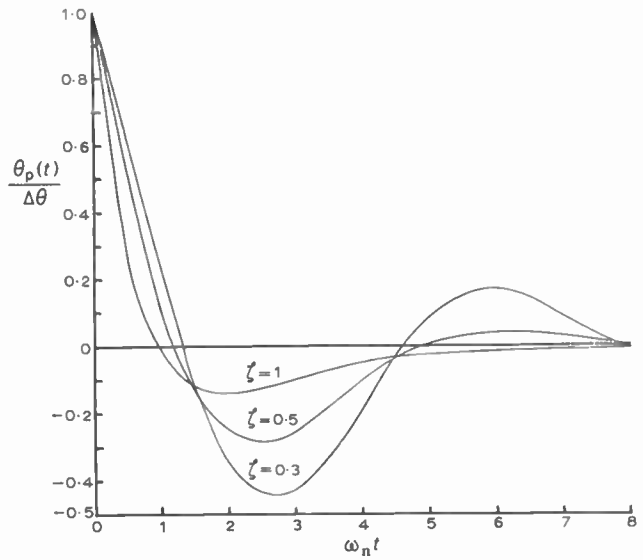
$$\Delta\theta = \theta_{us} - \theta_l \simeq \Delta\omega_p t_s/T_1(t_s/2+T_2), \quad \text{.....(15)}$$

and the loop responds to $\Delta\theta$ as if it were a step in phase at the system input.

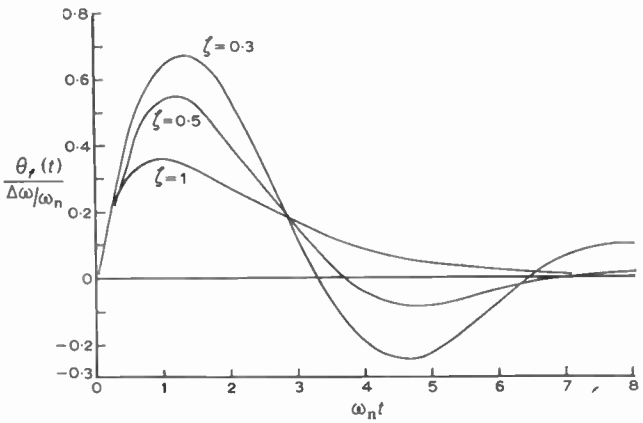
It can be seen from Fig. 10 and equation (12a) that the frequency error after t_s seconds is

$$\Delta\omega_p\left(\frac{t_s+T_2}{T_1}\right).$$

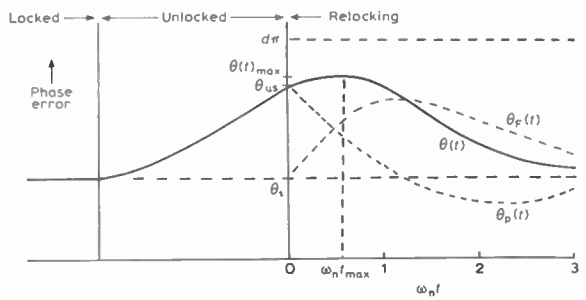
Upon reapplication of drive however the loop senses a frequency error dictated by the voltage across C only and not the reduced voltage due to R_1 and R_2 (see



(a) Phase error $\theta_p(t)$ due to a step in phase $\Delta\theta$.



(b) Phase error $\theta_f(t)$ due to a step in frequency $\Delta\omega$.



(c) The total relocking phase behaviour.

Fig. 12

Fig. 2). This leads to an effective frequency drift of

$$\Delta\omega = \Delta\omega_p t_s/T_1, \qquad \text{.....(16)}$$

and the loop responds as if a step of $\Delta\omega$ radians per second is applied to the input.

4.2.2. Evaluating the phase transient

To evaluate the total loop response upon relocking, the behaviour for each of the errors $\Delta\theta$, $\Delta\omega$ and θ_l must be considered separately. The total response, assuming a linear system, is then the sum of the individual responses. The instantaneous phase error for each of the step inputs $\Delta\theta$ and $\Delta\omega$ has been given[3]:

$$\theta_p(t) = \Delta\theta\left(\cos(\omega_n t\sqrt{1-\zeta^2})\frac{\zeta}{\sqrt{1-\zeta^2}} \times \right.$$
$$\left. \times \sin(\omega_n t\sqrt{1-\zeta^2})\right)\exp(1-\zeta\omega_n t) \quad \text{...(17)}$$

$$\theta_f(t) = \frac{\Delta\omega}{\omega_n}\left(\frac{1}{\sqrt{1-\zeta^2}}\sin(\omega_n t\sqrt{1-\zeta^2})\right) \times$$
$$\times \exp(-\zeta\omega_n t). \quad \text{...(18)}$$

Equations (17) and (18) are plotted in Fig. 12(a) and (b) and the general form of the total response in Fig. 12(c).

Figure 12(c) shows that the total phase error $\theta(t)$ can increase over the value θ_{us} to a maximum value $\theta(t)_{max}$ dependent on ζ; the maximum occurring between 0 and $1/\omega_n$ seconds after the initiation of relock. (Note that the maximum error can in fact be the initial error, θ_{us}.) Thus equation (14) is only a prerequisite for successful relocking and we see now from Fig. 12(c) that

$$\theta(t)_{max} < d\pi \qquad \text{.....(19)}$$

ensures relocking without a reduction in loop gain.

To evaluate $\theta(t)_{max}$ for a given set of loop constants, the equations (17) and (18) could be used, but this is tedious. Instead, it can be shown from these equations that the maximum error $\theta(t)_{max}$ occurs at a time $\omega_n t_{max}$ such that:

for $\zeta > 1$,

$$\tanh(\omega_n t_{max}\sqrt{\zeta^2-1}) = \sqrt{\zeta^2-1} \times$$
$$\times\left[\frac{2\zeta\dfrac{\Delta\theta}{\Delta\omega/\omega_n}-1}{2(\zeta^2-1)\dfrac{\Delta\theta}{\Delta\omega/\omega_n}-\zeta}\right], \qquad \text{.....(20a)}$$

for $\zeta = 1$,

$$\omega_n t_{max} = \frac{2\dfrac{\Delta\theta}{\Delta\omega/\omega_n}-1}{\dfrac{\Delta\theta}{\Delta\omega/\omega_n}-1}, \qquad \text{.....(20b)}$$

and for $\zeta < 1$,

$$\tan(\omega_n t_{max}\sqrt{1-\zeta^2}) = \sqrt{1-\zeta^2} \times$$
$$\times\left[\frac{2\zeta\dfrac{\Delta\theta}{\Delta\omega/\omega_n}-1}{2(\zeta^2-1)\dfrac{\Delta\theta}{\Delta\omega/\omega_n}-\zeta}\right] \qquad \text{.....(20c)}$$

Then Fig. 12 or equations (17) and (18) can be used to evaluate $\theta_p(t)_{max}$ and $\theta_f(t)_{max}$ at $\omega_n t_{max}$. The total phase error is seen from Fig. 12(c) to be

$$\theta(t)_{max} = \theta_l + \theta_p(t)_{max} + \theta_f(t)_{max} \qquad \text{.....(21)}$$

and equation (19) can be applied as a design check.

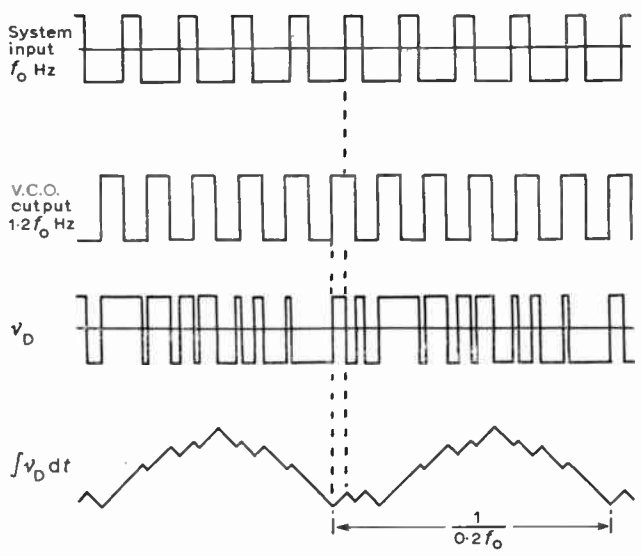


Fig. 13. The pull-in beat-note,

4.2.3. The effect of the phase transient on the raster

It can be seen from Fig. 12(c) that the v.c.o. phase 'rings' before approaching the static error after an interval of the order of $10/\omega_n$ seconds. If the phase is still settling when picture information commences, the top of the raster will be horizontally displaced. Thus we must ensure that ω_n and ζ are chosen so that the phase error is sufficiently small after the field blanking period.

5. Pull-in or Capture Range

5.1. *Pull-in with a Sinusoidal Input to the Loop*

It is possible for the unlocked loop to achieve lock when the input frequency is quite far removed from the v.c.o. frequency. The v.c.o. 'walks' towards the input frequency, the phase detector providing the necessary control voltage.

The output of the phase detector in the unlocked condition contains a component at the difference frequency between the v.c.o. and the loop input. It has been described[3,5–9] how this beat-note has a d.c. component and will tend to pull the v.c.o. towards lock. For a sinusoidal system input, the maximum initial frequency difference from which the loop will lock itself, called the pull-in frequency ω_p, has been given[2,3,7]:

$$\Delta\omega_p = \sqrt{2}(2\zeta\omega_n K_O K_D - \omega_n^2)^{\frac{1}{2}}. \quad \text{.....(22a)}$$

This holds, providing

$$\omega_n < 0{\cdot}4\, K_O K_D$$

and for a high gain loop becomes

$$\Delta\omega_p = 2\sqrt{\zeta\omega_n K_O K_D}. \quad \text{.....(22b)}$$

Associated with pull-in range is a time to achieve lock, called the pull-in time, but in this application is of no consequence since it will always be of the order of milliseconds.

5.2. *Pull-in with a Pulse Input to the Loop*

In Section 2.2, the nature of the phase detector output for a pulse drive was considered with the loop in the locked condition. Now consider Fig. 3 with the v.c.o frequency different from the input frequency. The waveform v_D becomes a complex digital signal which is difficult to analyse. Some insight into the nature of the beat-note waveform can be gained by considering the phase detector output with the v.c.o. frequency 20% higher than the input frequency. Further, assume the output to be perfectly integrated by the filter of Fig. 2. This is shown in Fig. 13.

The filter output is seen to be a complex wave with a large component at $1{\cdot}2f_0 - f_0$ or $0{\cdot}2f_0$, the beat frequency. This crude analysis does not consider the fact that the v.c.o. actually becomes frequency modulated by the beat-note. (It is in fact this frequency modulation which causes the beat-note to have a d.c. component and makes pull-in possible.) Further reasoning shows that the beat-note amplitude and waveshape depends on both drive pulse width and frequency difference. For what follows it is intuitively assumed that the d.c. component of the beat-note waveform is proportional to the drive pulse width. Thus the d.c. component feeding the v.c.o. depends on the duty cycle, and this is as if the phase detector sensitivity K_D is a function of d. Assuming that when $d = 0{\cdot}5$ (a square wave), the effective loop gain during pull-in has the same value as for a sinusoidally driven loop, we modify equation (22b) to:

$$\Delta\omega_p = 2\sqrt{\zeta\omega_n 2d K_O K_D}. \quad \text{.....(23)}$$

5.3. *Pull-in Experiments*

In an attempt to prove the reasoning presented in Section 5.2, a series of experiments was performed. The loop was fed from a pulse generator, and the maximum v.c.o. frequency error $\Delta\omega_p$ from which pull-in occurred was plotted against pulse width. When the v.c.o. was running above the input frequency (15·625 kHz), pull-in was achieved from a frequency error higher than for the v.c.o. running below the input frequency. This is because the oscillator sensitivity is proportional to the free-running frequency ω_p giving an effective sensitivity of

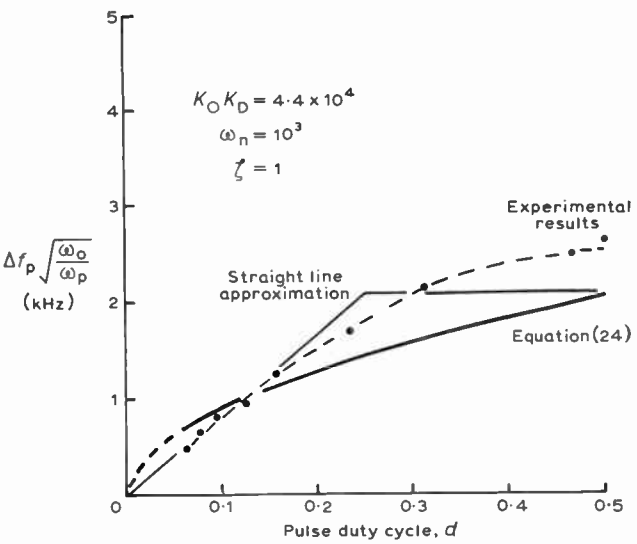


Fig. 14. Pull-in performance.

ω_p/ω_0 K_O radians per second per volt. Equation (23) can then be written:

$$\Delta f_p \sqrt{\frac{\omega_0}{\omega_p}} = \frac{2}{2\pi}\sqrt{\zeta\omega_n 2d K_O K_D}. \quad \text{.....(24)}$$

Equation (24) is plotted in Fig. 14 together with the experimental results for $\omega_n = 10^3$, $\zeta = 1$, $K_O K_D = 4{\cdot}4 \times 10^4$. The measured values for Δf_p were weighted by $\sqrt{\frac{\omega_0}{\omega_p}}$ then the mean of the weighted positive and negative values were plotted.

Figure 14 shows that for $d = 0{\cdot}5$, i.e. a square wave drive, the pull-in performance is better than predicted by equation (24). This seems to indicate that the d.c. component of the beat-note is higher than it would be for a sinusoidal system input. It is reasonable to expect this, considering the improved phase detector linearity indicated in Fig. 5. Conversely, at low values of d (narrow pulse drive), the pull-in performance is worse than predicted. Note that equation (22a) holds providing $\omega_n < 0{\cdot}4\, K_O K_D$ and for this discussion, $K_O K_D$ is taken to mean the effective value, $2dK_O K_D$. Thus it seems likely that experimental data would not show a good correlation with equation (24) at low duty cycles.

5.4. *Intuitive Design Assumption*

Figure 14 shows that if equation (24) is correct, small values of d (about 0·1) would achieve pull-in values of the order of $\frac{1}{2}$ of that for a square wave ($d = 0{\cdot}5$). In an attempt to clarify this unlikely state of affairs, more experiments were performed with different values of loop constants. It was then found that a straight line seemed a reasonable fit to the curves for d from 0 to 0·25; the frequency value at $d = 0{\cdot}25$ being that calculated from equation (22a). This is shown in Fig. 14 for comparison with equation (24) and in Fig. 15 for a few loop designs.

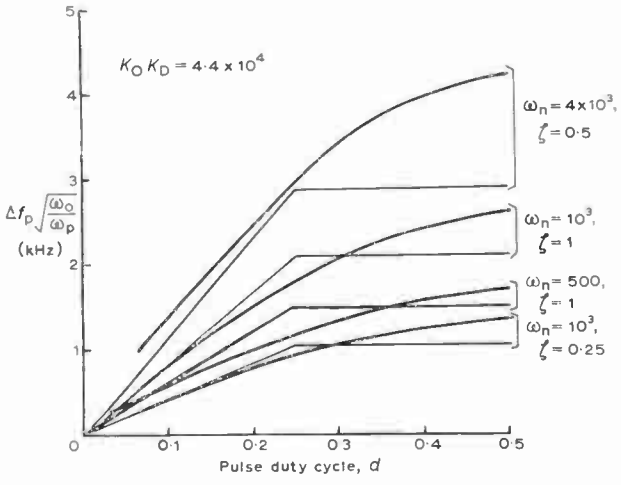


Fig. 15. Experimental data compared with pull-in approximation.

As a reasonable design equation we then use, for

$$0 < d < 0{\cdot}25$$

$$\Delta\omega_p \sqrt{\frac{\omega_0}{\omega_p}} = 2 \times 4d\sqrt{\zeta\omega_n K_0 K_D}$$

And for small pull-in values:

$$\Delta\omega_p \simeq 8d\sqrt{\zeta\omega_n K_O K_D}. \quad \text{.....(25)}$$

6. Design Example

In general, p.l.l. design is always a compromise, and no single optimum exists for all the desired properties. A sample design is given here to show how the performance criteria presented throughout the paper are interrelated. The design follows broadly the following sequence:

specify $\Delta\omega_p$, d, $K_O K_D$
decide upon a value for ζ
calculate ω_n, considering pull-in performance
calculate T_1 and T_2, and design the filter
check for satisfactory relock after field sync.
check the effect of the relock transient on the picture
assess noise performance experimentally

6.1. *System Specification*

Using an LM565 i.c. for a c.c.t.v. application, a design having the following specification has been done.

television system	625 lines, 50 fields/s, 2 : 1 inter-lace
line sync. pulses	5 μs
duty cycle, d	5/64 = 0·078
field sync.	One broad pulse occupying 7 lines or 0·448 ms
field blanking period	20 lines or 1·28 ms
pull-in performance	Allow ±4% v.c.o. tuning error or ±625 Hz error (centre frequency = 15·625 kHz)
noise performance	No perceptible line jitter at 20 dB black to white video to r.m.s. noise (white Gaussian noise, 100 kHz bandwidth)
loop gain[1], $K_O K_D$	$4{\cdot}4 \times 10^4\ s^{-1}$ (at ±6 V supply to the i.c.)

6.2. *The Design*

Before the design proceeds, a value for ζ must be decided upon. $\zeta = 0{\cdot}5$, for minimum noise bandwidth, is chosen for this case, but in many designs leads to an excessive phase transient after field sync. ($\zeta = 1$ is a common choice).

(i) Calculate ω_n

Equation (25) gives:

$$\omega_n = 1{\cdot}79 \times 10^3 \text{ rad/s (285 Hz)}$$

(ii) Calculate T_1 and T_2

Equations (9) and (10) (or Fig. 7) give:

$$T_1 = 13{\cdot}2 \text{ ms}$$
$$T_2 = 0{\cdot}54 \text{ ms}$$

(iii) The filter

The LM565 Application Note[1] gives R_1 in Fig. 2 as:

$$R_1 = 3{\cdot}6\ \text{k}\Omega$$

Equations (1) and (2) then yield:

$$C = 3{\cdot}65\ \mu\text{F}$$

$$R_2 = 148\ \Omega$$

(iv) Static phase error

Equation (7) gives

$$\theta_l = 0{\cdot}086 \text{ or } -0{\cdot}093 \text{ rad}$$

i.e. for the v.c.o. tuned to $\omega_0+\Delta\omega_p$, the raster is displaced 1·4% of picture width to the right (triggering early); and for the v.c.o. running low, the displacement is 1·5% to the left. Take θ_i as the mean of the above values; i.e.

$$\theta_l \simeq 0{\cdot}09 \text{ rad}$$

for all following calculations. (This can also be expressed by equation (4) as a timing error t_l of 0·9 μs.)

(v) The v.c.o. frequency drift during field sync.

Equation (12a) gives the frequency after 0·448 ms of no locking information as:

$$\omega_i = 2\pi \times 15{\cdot}672 \times 10^3 \text{ rad/s}$$

i.e. a gain of 47 Hz over the input during field sync.

(vi) The phase advance during field sync.

Equation (15) gives the phase advance corresponding to the 47 Hz frequency drift above as:

$$\Delta\theta = 0{\cdot}1 \text{ rad}$$

This gives a total phase error prior to relock, from equation (13a) of

$$\theta_{us} = 0{\cdot}19 \text{ rad}$$

or a timing error (see Fig. 11) of:

$$t_{us} = 1{\cdot}94\ \mu\text{s}$$

The initial requirement for relock, equation (14), is satisfied since

$$d\pi \simeq 0{\cdot}25 > 0{\cdot}19$$

(vii) The effective frequency step

Equation (16) gives:

$$\Delta\omega = 2\pi \times 20{\cdot}6 \text{ rad/s}$$

i.e. the loop is required to relock from an effective 20·6 Hz error, and not the actual 47 Hz drift calculated in (v) above.

(viii) The total relocking transient

Equation (20c) gives:

$$\tan 0{\cdot}87\,\omega_n t_{max} = -0{\cdot}302$$

This shows that the first maximum for $\theta(t)$ is in fact at

$$\omega_n t = 0$$

giving a value of

$$\theta(t)_{max} = \theta_{us} = 0{\cdot}19 \text{ rad}$$

Equation (20c) can also give us the second maximum:

$$0{\cdot}87\,\omega_n t_{max} = \pi - 0{\cdot}29$$

Therefore,

$$\omega_n t_{max} = 3{\cdot}29$$

Figure 12 or equations (17) and (18) yield:

$$\theta_p(t)_{max} = -0{\cdot}022 \text{ rad},$$

and

$$\theta_f(t)_{max} = 0{\cdot}0051 \text{ rad}$$

Equation (21) gives the total phase error as:

$$\theta(t)_{max} = 0{\cdot}073 \text{ rad}$$

(ix) Horizontal raster shift

To assess the effect of relocking on the picture, a few calculations of the picture shift are given in Table 1. The v.c.o. is assumed to be set to $\omega_0+\Delta\omega_p$ or 16·25 kHz. The

Table 1. Raster shift due to relock transient

Position	Shift	Remarks
Top of raster	0·4% right	$\theta_p(t)+\theta_f(t)$ at $t = 0{\cdot}83$ ms (13 lines after initiation of relock)
5·4% of picture height from top	0·26% left	Second phase maximum at $\omega_n t = 3{\cdot}29$ (see (viii) above)
10% of picture height from top	0·09% left	
20% of picture height from top	<0·01% right	loop can be considered settled for $\omega_n t > 8$; see Fig. 12

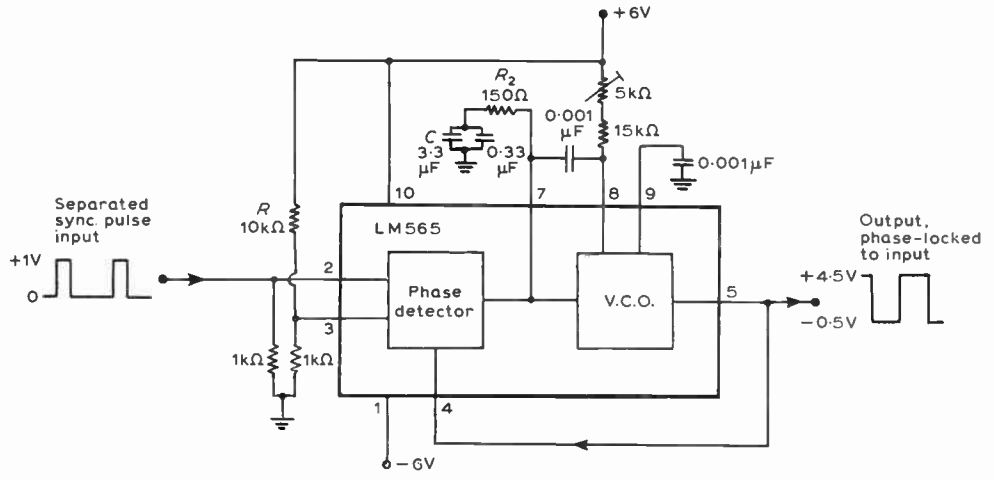


Fig. 16.
Complete flywheel generator.

shift is expressed relative to the static shift (see (iv) above) of 1·4% to the right.

6.3. *Experimental Check of the Design*

The design proposed in Section 6.2 was implemented as shown in Fig. 16.

The component values are all as given or calculated by the application note,[1,2] with the exception of R. This resistor biases pin 3 to about 0·5 V to enable the loop to cope with the digital input signal indicated. Referring to Fig. 9, this gives an input threshold voltage of 0·5 V, the loop limiting[2] 10 mV beyond this, i.e. an amplitude 'window' of 10 mV.

The pull-in performance was measured, giving a pull-in frequency Δf_p of 730 Hz.

The noise performance was assessed by injecting 0·1 V r.m.s. of 100 kHz bandwidth white Gaussian noise into pin 2 with the loop locked. The output phase jitter was not perceptible, satisfying the specification given in Section 6.1.

The relocking transient was found to be subjectively acceptable.

7. Conclusions

A design procedure has been evolved, which gives a satisfactory, if not precise, solution to an engineering problem namely, that of designing the flywheel sync. circuit using 'state of the art' techniques. Although the paper deals specifically with the p.l.l. in a television application, many of the principles discussed are applicable to any digitally-driven p.l.l. system.

The author has used the p.l.l. design method given in developing a line drive unit and a clock pulse generator for a divide-by-625 circuit which were both for use in a specialized educational television system.

Plans are underway at the time of writing to generalize, with the aid of a computer, p.l.l. design to cover other uses such as colour sub-carrier regeneration.

8. Acknowledgments

The author is grateful to the Director of the Natal College for Advanced Technical Education for permission to publish this paper.

9. References

1. 'LM565/LM565C Phase Locked Loops', Application Note, National Semiconductor Corp., 1971.
2. Mills, T. B., 'The Phase Locked Loop I.C. as a Communication System Building Block', Application Note AN46, National Semiconductor Corp., 1971.
3. Gardiner, F. M., 'Phaselock Techniques', (Wiley, New York, 1966).
4. Roulston, J. F., 'Applying the phase-locked loop in communications and instrumentation', *The Radio and Electronic Engineer*, 41, pp. 315–20, July 1971.
5. Carnt, P. S. and Townsend, G. B., 'Colour Television, The N.T.S.C. System Principles and Practice', Vol. 1, (Iliffe, London, 1961).
6. Richman, D., 'Colour-carrier reference phase synchronization, accuracy in N.T.S.C. color television', *Proc. I.R.E.*, 42, pp. 106–33, January 1954.
7. Gruen, W. J., 'Theory of a.f.c. synchronization', *Proc. Inst. Radio Engrs*, 41, pp. 1043–8, August 1953.
8. Rey, T. J., 'Automatic phase control: theory and design', *Proc. I.R.E.*, 48, pp. 1760–71, October 1960.
9. McAleer, H. T., 'A new look at the phase-locked oscillator', *Proc. I.R.E.*, 47, pp. 1137–43, June 1959.

Manuscript received by the Institution on 1st May 1972. (Paper No. 1480/Com. 56.)

The Author

Mr. P. Pomeroy (Graduate 1970) obtained a Technician's Diploma in Electronics at the Witwatersrand College for Advanced Technical Education in 1966 whilst employed by the Electricity Supply Commission. Following part-time study while working on acoustics research at the South African Railway Laboratory, he obtained Graduateship of the SAIEE in 1969 and of the IERE in 1970. Since 1969 he has been a lecturer in electronics and television at the Natal College for Advanced Technical Education and has been responsible for developing several items of specialized educational equipment.

U.D.C. 621.372.5

Digital Filters: A Template Method of Design

E. R. BROAD, B.A.*
and
P. F. ADAMS, B.Sc.†

Presented at the IERE Conference on Digital Processing of Signals in Communications held in Loughborough from 11th to 13th April 1972.

SUMMARY

It is shown that the frequency response of digital filters of the cascade biquadratic (ratio of two quadratic polynomials) form can be described by a set of templates. These templates enable the effects of coefficient quantization to be seen. Consequently, by using the templates to design filters, coefficient quantization can be taken into account in the initial stages of the designing process. Examples of filter designs are given.

* *Formerly with the Post Office Research Department.*
† *Post Office Research Department, Dollis Hill, London, NW2 7DT.*

1. Introduction

The mathematical procedures for the design of digital filters which have been developed so far, while straightforward, do not show directly the influence of the multiplier coefficients on the attenuation/frequency characteristic of the filter. This can be important when coefficient accuracy is limited in real hardware. The template method outlined in this paper permits the influence of the coefficients to be seen more readily and provides a design method for digital filters akin to procedures already familiar in analogue filter design.

The method is concerned only with the discrimination behaviour of the digital filter. It assumes the use of the cascade connexion of a number of biquadratic sections or, rather, of a number of quadratic sections which may be either recursive or non-recursive. This form has some advantages both theoretically and practically over others and is dealt with adequately in the literature.[1, 2] The first-order section can be regarded as a 'half-section' or degenerate quadratic section. The discrimination/frequency characteristic of a quadratic section conforms to a set of well defined templates when the frequency scale is changed from frequency f to $\cos(2\pi f/f_s)$ where f_s is the sampling rate. These templates can be used as in other template methods[3] to build up a required discrimination characteristic.

2. Discrimination Templates

In this section it is shown how the discrimination characteristic of a digital filter can be described by a set of templates and that, using the templates, filters can be designed taking into account coefficient quantization.

2.1. *The Discrimination of Quadratic Sections*

The pulse transfer function

$$G(z) = 1 + Az^{-1} + Bz^{-2}$$

has an amplitude gain/frequency characteristic $|G(j\omega)|$ given by (see Appendix)

$$|G(j\omega)| = [A^2 + (1-B)^2 + 2A(1+B)\cos\theta + 4B\cos^2\theta]^{\frac{1}{2}}. \quad \ldots\ldots(1)$$

where

$$\theta = 2\pi f/f_s = \omega T$$
$$T = 1/f_s$$

is the sampling period.

In terms of decibels this becomes

$$G = 20\log_{10}(|G(j\omega)|)$$

i.e.

$$G = 10\log_{10}[4B\cos^2\theta + 2A(1+B)\cos\theta + A^2 + (1-B)^2]. \quad \ldots\ldots(2)$$

It is convenient to work in terms of attenuation R dB rather than gain, so putting $R = -G$ and rearranging the expression

$$R = -10\log_{10}[4(\cos\theta - C)^2 + D] - 10\log_{10}(B) \quad \ldots\ldots(3)$$

where

$$C = -A(1+B)/4B$$

and

$$D = (4B - A^2)(1-B)^2/4B^2.$$

The second term of eqn. (3) is independent of θ and is constant for a given B. The first term is a function of θ and determines the shape of the attenuation/θ characteristic. This 'shape' term, which will be designated P, is useful on its own because the absolute values of attenuation can be altered by any frequency-independent gain.

Therefore,

$$P = -10 \log_{10} (4(\cos\theta - C)^2 + D). \quad \ldots\ldots(4)$$

Inspection of eqn. (4) shows that the shape of the P characteristic plotted against $\cos\theta$ is independent of C. A change in C only moves the whole characteristic parallel to the $\cos\theta$ axis.

Thus putting $y = \cos\theta - C$ in the expression for P

$$P = -10 \log_{10} (4y^2 + D). \quad \ldots\ldots(5)$$

$P(y)$ gives a family of curves which, by suitable positioning along the $\cos\theta$ axis, can describe the $P/\cos\theta$ characteristic of all digital filters having a pulse transfer function $G(z)$ of the non-recursive quadratic form. It should be noted that although y can take any real value it is only the portion of the curve appearing in the range $1 \geqslant \cos\theta \geqslant -1$ that is valid.

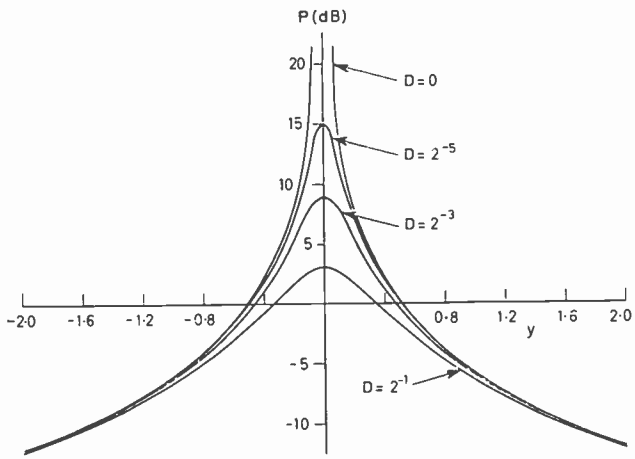


Fig. 1. Discrimination templates.

For the pulse transfer function

$$G_r(z) = \frac{1}{1 + Az^{-1} + Bz^{-2}}$$

it can be shown by similar reasoning that one of a family of curves defined by the expression

$$P_r = -P = 10 \log_{10} (4y^2 + D) \quad \ldots\ldots(6)$$

suitably positioned on the $\cos\theta$ axis can describe the $P_r/\cos\theta$ characteristic of any $G_r(z)$.

P and P_r are characteristics which give a particular representation of the discrimination characteristic of a quadratic section. It should be noted, however, that whereas P and P_r give absolute values, absolute values are not necessary for the discrimination characteristic. Hence in using the curves in the discrimination/$\cos\theta$ plane translation along the discrimination axis is permissible.

2.2. *The Discrimination of First-order Sections*

The first-order section

$$G(z) = 1 + Az^{-1}$$

has an amplitude gain/frequency characteristic given by

$$|G(j\omega)| = (1 + A^2 + 2A\cos\theta)^{\frac{1}{2}}. \quad \ldots\ldots(7)$$

In terms of attenuation R dB this becomes

$$R = -10 \log_{10} [2\cos\theta + (1 + A^2)/A] - 10 \log_{10} (A). \quad \ldots\ldots(8)$$

With $C_L = -(1 + A^2)/2A$ and calling, as before, the 'shape' term P, then

$$P = -10 \log_{10} (2(\cos\theta - C_L)). \quad \ldots\ldots(9)$$

Putting $y = \cos\theta - C_L$ the $P/\cos\theta$ characteristic of a first-order section can be described by a single curve

$$P = -10 \log_{10} (2y) \quad \ldots\ldots(10)$$

suitably positioned along the $\cos\theta$ axis. It will be noticed that eqn. (10) gives P values of one-half those obtained from eqn. (5) with $D = 0$.

2.3. *The Templates*

Equation (5) describes a family of curves which, with due attention to the sign of P can give the shape of the discrimination/$\cos\theta$ characteristic of any quadratic section, recursive or non-recursive. These curves need be moved only along the $\cos\theta$ axis because translation along the discrimination axis represents a change in frequency-independent gain only and does not affect discrimination. However, translation along the discrimination axis is possible and sometimes desirable. Furthermore, the combined discrimination/$\cos\theta$ characteristic of two or more cascaded quadratic sections is formed from the algebraic sum of the individual discrimination/$\cos\theta$ characteristics. These properties mean that the expression

$$P = -10 \log_{10} (4y^2 + D)$$

can define a set of templates which can be used to build up the discrimination/$\cos\theta$ characteristic of any cascaded biquadratic filter.

A set of templates for $D = 2^{-n}$ where $n = 1$, 3, 5 and $D = 0$ is shown in Fig. 1. It is possible for D to have negative values but such templates have a restricted use, are associated with larger coefficients (longer coefficient word lengths) and have no advantages when used in filter design. Table 1 gives the values of a larger range of templates.

The procedure to obtain the discrimination/$\cos\theta$ characteristic of a filter of the form

$$G(z) = \frac{(1 + A_1 z^{-1} + B_1 z^{-2})(1 + A_2 z^{-1} + B_2 z^{-2})(\ldots)\ldots}{(1 + A_3 z^{-1} + B_3 z^{-2})(1 + A_4 z^{-1} + B_4 z^{-2})(\ldots)\ldots}$$

is to choose the correctly shaped template, i.e. the one of D value given by $D_i = (4B_i - A_i^2)(1 - B_i)^2/4B_i^2$ for each quadratic section and position it on the $\cos\theta$ axis with its centre line of symmetry passing through the point $\cos\theta = C_i = -A_i(1 + B_i)/4B_i$ and parallel to the discrimination axis. Thus the discrimination/$\cos\theta$ characteristic of each quadratic section is obtained. If first-order sections are used, templates of D value zero are

Table 1

P (dB) for given values of y and D

	$D=2^{-n}$									
n	1	2	3	4	5	6	7	8	9	∞
y										
0·0	3·010	6·021	9·031	12·041	15·051	18·062	21·072	24·082	27·093	∞
0·1	2·676	5·376	7·825	9·893	11·472	12·547	13·205	13·575	13·772	13·979
0·2	1·805	3·872	5·452	6·527	7·184	7·554	7·752	7·854	7·906	7·959
0·3	0·655	2·147	3·143	3·742	4·076	4·253	4·344	4·390	4·414	4·437
0·4	−0·569	0·506	1·163	1·534	1·731	1·833	1·886	1·912	1·925	1·938
0·5	−1·761	−0·969	−0·512	−0·263	−0·134	−0·067	−0·034	−0·017	−0·008	0·000
0·6	−2·878	−2·279	−1·945	−1·768	−1·677	−1·631	−1·607	−1·595	−1·590	−1·584
0·7	−3·909	−3·444	−3·191	−3·059	−2·991	−2·957	−2·940	−2·931	−2·927	−2·923
0·8	−4·857	−4·487	−4·289	−4·187	−4·135	−4·109	−4·096	−4·089	−4·086	−4·082
0·9	−5·729	−5·428	−5·270	−5·188	−5·147	−5·126	−5·116	−5·111	−5·108	−5·106
1·0	−6·532	−6·284	−6·154	−6·088	−6·054	−6·038	−6·029	−6·025	−6·023	−6·012
1·1	−7·275	−7·067	−6·959	−6·904	−6·876	−6·863	−6·856	−6·852	−6·850	−6·849
1·2	−7·966	−7·789	−7·698	−7·651	−7·628	−7·616	−7·610	−7·607	−7·606	−7·604
1·3	−8·609	−8·457	−8·379	−8·339	−8·320	−8·310	−8·305	−8·302	−8·301	−8·300
1·4	−9·212	−9·080	−9·012	−8·978	−8·960	−8·952	−8·948	−8·945	−8·944	−8·943
1·5	−9·777	−9·661	−9·602	−9·573	−9·558	−9·550	−9·546	−9·544	−9·543	−9·542
1·6	−10·310	−10·208	−10·156	−10·129	−10·116	−10·110	−10·106	−10·105	−10·104	−10·103
1·7	−10·813	−10·722	−10·676	−10·653	−10·641	−10·635	−10·633	−10·631	−10·630	−10·630
1·8	−11·290	−11·209	−11·168	−11·147	−11·137	−11·131	−11·129	−11·127	−11·127	−11·126
1·9	−11·744	−11·670	−11·633	−11·614	−11·605	−11·600	−11·598	−11·597	−11·596	−11·596
2·0	−12·175	−12·109	−12·075	−12·058	−12·050	−12·045	−12·043	−12·042	−12·042	−12·041

positioned on the $\cos\theta$ scale at $C_i = -(1+A_i^2)/2A_i$. The characteristic of each section is then obtained by halving the discrimination values given by the templates. (Thus the first-order section can be thought of as a 'half-section' and its template as a 'half-template'.) The discrimination/$\cos\theta$ characteristic of the whole filter is then obtained by adding the characteristics of the non-recursive sections and, since $P_R = -P$ (see eqn. (6)) subtracting the characteristics of the recursive sections.

3. Use of the Templates

It has been shown that the discrimination/$\cos\theta$ characteristic of any digital filter expressed in the cascaded biquadratic form can be built up from a set of templates; but since every quadratic section has a unique template placed at a unique position which gives its discrimination/$\cos\theta$ characteristic, the converse is also true. A template placed in any position will give a discrimination/$\cos\theta$ characteristic that has an associated quadratic section. In the case of first-order sections, the placing of a 'half template' is restricted to the centre line of symmetry not being placed inside the range $-1 < \cos\theta < 1$, i.e. $|C_L| \geqslant 1$. This is because coefficients are restricted to being real.

Thus, to design a filter to a given discrimination/frequency specification the specification is first translated to the discrimination/$\cos\theta$ plane. Templates are then used to build up a characteristic that meets the specification. The D and C values associated with each template and its position on the $\cos\theta$ axis are noted. From these, the values of the coefficients A and B for each quadratic are obtained.

This in itself is a useful design procedure but it is possible to take into account coefficient quantization.

4. The Effects of Coefficient Quantization

The restriction of coefficient word lengths in real hardware means that instead of a continuous range of values of A and B there are quantized steps. Since C and D are functions of A and B, i.e.

$$C = -A(1+B)/4B$$

and

$$D = (4B-A^2)(1-B)^2/4B^2$$

only certain templates in certain positions on the $\cos\theta$ axis are allowed. Hence to take coefficient quantization into account in the design of filters a knowledge of the allowed templates and their positions is required for a given coefficient word length. So far, the ranges of values of A and B have not been restricted. It has been found however that the ranges $2 \geqslant A \geqslant -2$ and $1 \geqslant B > 0$ give a useful selection of templates. These ranges can give negative D values if $A^2 > 4B$ so a further restriction is $4B \geqslant A^2$. With these restrictions there will now be a finite number of allowed D and C values. The best way of showing these is to plot them in the $C-D$ plane. It is only necessary to plot points corresponding to positive A coefficients because changing the sign of A does not change the value of D and only changes the sign of C.

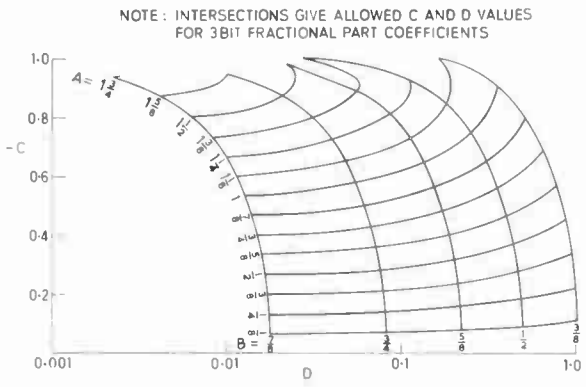


Fig. 2. Lines of constant A and B.

A plot for binary coefficients with three bit fractional parts is shown in Fig. 2. For clarity, lines of constant A and B are plotted but it is the intersections of these lines that give the allowed D and C values. More accurate values than those on the graph are given in Table 2. Not shown on the graph is the special case of $B = 1$. For this case, the effect of the quantization of A is easily seen. Putting $B = 1$ in the formula for C gives $C = -A/2$. Thus the template with $D = 0$ can be placed only at positions C on the $\cos\theta$ axis that are integer multiples of $2^{-(n+1)}$ where n is the number of fractional bits used in the binary A coefficient. It must be noted that templates with $D = 0$ can be used only for non-recursive sections; recursive sections are unstable when $B = 1$.

In designing a filter only the allowed templates and the allowed template positions are used thus taking into account coefficient quantization during the filter design.

5. The Design of Filters using the Templates

The previous sections have shown how the behaviour of the cascaded biquadratic form of digital filter can be described in terms of a set of templates and that the effects

Table 2

Values of C and D for coefficients A and B with 3-bit fractional parts

B	$\frac{1}{4}$		$\frac{3}{8}$		$\frac{1}{2}$	
A	C	D	C	D	C	D
0	0·0000	2·5000	0·0000	1·0417	0·0000	0·5000
0·125	0·1563	2·2148	0·1145	1·0308	0·0938	0·4961
0·25	0·3125	2·1094	0·2292	0·9983	0·1875	0·4844
0·375	0·4688	1·9336	0·3437	0·9440	0·2813	0·4648
0·5	0.6250	1·6875	0·4583	0·8681	0·3750	0·4375
0·625	0·7813	1·3711	0·5729	0·7704	0·4688	0·4023
0·75	0·9375	0·9844	0·6875	0·6510	0·5625	0·3594
0·875	1·0938	0·5273	0·8021	0·5100	0·6563	0·3086
1	1·2500	0·000	0·9167	0·3472	0·7500	0·2500
1·125	—	—	1·0313	0·1628	0·8438	0·1836
1·25	—	—	—	—	0·9375	0·1094
1·375	—	—	—	—	1·0313	0·02734

B	$\frac{5}{8}$		$\frac{3}{4}$		$\frac{7}{8}$	
A	C	D	C	D	C	D
0	0·0000	0·2250	0·0000	0·08333	0·0000	0·01786
0·125	0·0813	0·2236	0·0729	0·08290	0·0700	0·01778
0·25	0·1625	0·2194	0·1458	0·08160	0·1339	0·01754
0·375	0·2438	0·2123	0·2188	0·07943	0·2009	0·01714
0·5	0·3250	0·2025	0·2917	0·07639	0·2679	0·01658
0·625	0·4063	0·1898	0·3646	0·07248	0·3348	0·01586
0·75	0·4875	0·1744	0·4375	0·06771	0·4018	0·01499
0·875	0·5687	0·1561	0·5104	0·06207	0·4687	0·01395
1	0·6500	0·1350	0·5833	0·5556	0·5357	0·01276
1·125	0·7313	0·1111	0·6563	0·04818	0·6027	0·01140
1·25	0·8125	0·08438	0·7292	0·03993	0·6696	0·009885
1·375	0·8937	0·05484	0·8021	0·03082	0·7366	0·008211
1·5	0·9750	0·02250	0·8750	0·02083	0·8036	0·006378
1·625	—	—	0·9479	0·009983	0·8705	0·004385
1·75	—	—	—	—	0·9375	0·002232

of coefficient quantization can be taken into account when using the templates for filter design. However, if this approach is to be more than just interesting and enlightening, it must be able to produce filter designs that are worthwhile. There are many ways of using the templates to achieve a particular design specification. One fairly general method that has been used successfully is dealt with subsequently. It is instructive, however, to consider two general guidelines to the use of the templates.

As a non-recursive section has a peak of attenuation at the centre of its template, is should, in general, be used only with the centre of its template in the stop-band of any filter. Similarly, a recursive section should be used with its peak of gain in the pass-band.

For a given number of sections sharper cut-off can be obtained at the expense of pass band flatness, by using templates with small values of D, i.e. $B \rightarrow 1$. Conversely a flat pass-band is more easily obtained for a given number of sections by using templates with larger D values, sharpness of cut-off being sacrificed.

5.1 *A Design Method for Low- and High-Pass Filters*

The design of low- and high-pass filters is essentially the same because of the symmetry of the templates and the symmetry of the cos θ scale (ignoring sign) about its centre. The following description is in terms of low-pass filter design.

Many filter requirements fall into the category that can be defined in terms of a pass-band edge, frequency f_1, a stop-band edge, frequency f_2, a maximum pass-band tolerance p_{max} and a minimum pass-band to stop-band discrimination $p_{min}-p_{max}$ (Fig. 3(a)). The method of using the templates to meet this sort of specification exploits the fact that the stop-band and pass-band requirements are different. The difference is that the stop-band has only a single bound, i.e. p_{min} and the pass-band has two bounds, i.e. the frequency axis and p_{max}. Essentially the method is to use templates with $D = 0$ corresponding to non-recursive sections to obtain the required p_{min} and then to use templates corresponding to recursive sections to compensate the characteristic in the pass-band to achieve the pass-band tolerance.

To design a low-pass filter the procedure is as follows:

(i) It is assumed that p_{min}, p_{max}, f_1, f_2, the sampling rate f_s and the number n of fractional bits of the coefficients are chosen. The first step is to convert f_1 and f_2 to E_1 and E_2 respectively on the cos θ scale by $E_i = \cos(2\pi f_i/f_s)$ and draw the specification on the discrimination/cos θ graph (Fig. 3(b)).

(ii) A guess, based on previous experience, is made of the number of recursive sections that will be needed to compensate the characteristic in the pass-band. The recursive sections will affect the characteristic in the stop-band, so it is useful to make a pre-adjustment to the stop-band specification. Consideration of Fig. 1 shows that, at some distance from the centres of the templates, the slope dP/dy of the templates approximates

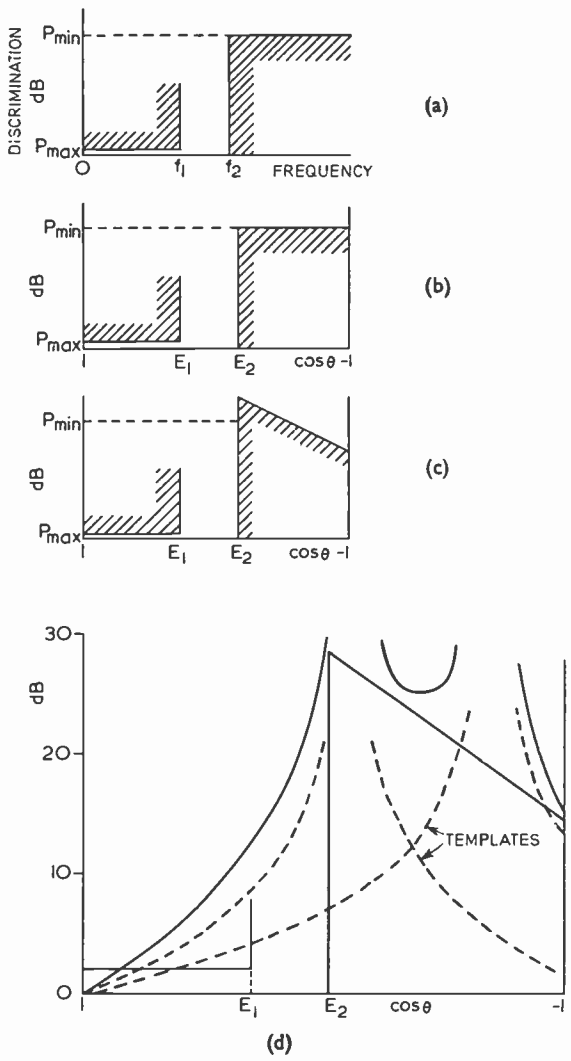


Fig. 3. Low-pass filter design.

± 6 dB/unit y. Thus a useful way of pre-adjusting the stop-band requirement is to add to it a line of slope 6 passing through the point $((2C_i - 1), 0)$, where C_i is the expected position of the ith recursive section template, for each recursive section (Fig. 3(c)).

(iii) Non-recursive section templates with D values of zero are now positioned in the stop-band so that added together they meet the adjusted stop-band requirement. (Fig. 3(d)). As the coefficients are allowed to have only n bit fractional parts and $C = -2A$ for $B = 1$ the template positions C are limited to integer multiples of $2^{-(n+1)}$ on the cos θ scale.

(iv) The filter characteristic in the pass-band has the general shape shown in Fig. 3(d). Templates corresponding to recursive sections are now used to bring this characteristic to within the tolerance p_{max}. However, as only certain templates are allowed because of the quantization of the coefficients, a chart similar to Fig. 2 only plotted for steps of 2^{-n} in A and B is used to select suitable templates.

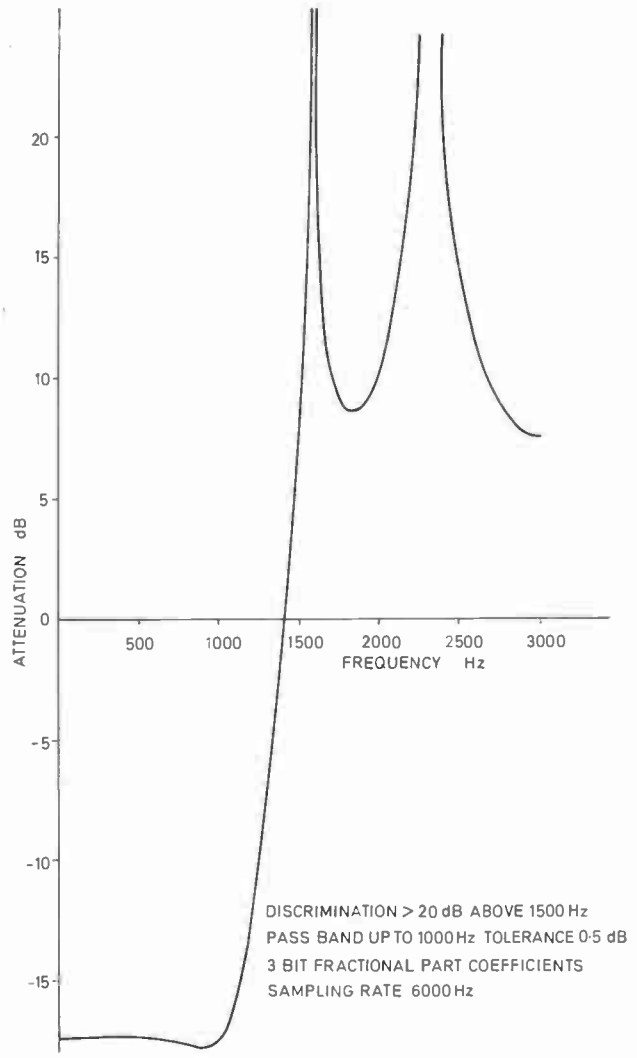


Fig. 4. Frequency response of a low-pass filter.

The method outlined above can be used to design a wide variety of filters but there are some difficulties that arise. As the filter specification becomes more stringent, the number of sections required to meet it increases. This makes the selection of templates and template positions more difficult. Also, more stringent specifications generally require more accurate coefficient quantization. A further problem is that, for constant $p_{\min}$ and $p_{\max}$ and a given width of transition region, as the transition region approaches the edges of the Nyquist band, more sections are required to meet the specifications and the choice of templates becomes more difficult.

Even with these difficulties this method has been used to design filters with quite stringent requirements. Figures 4 and 5 give the frequency responses of two filters designed in this way. Figure 4 shows the frequency response of a filter with the pulse transfer function

$$G(z) = \frac{(1+\frac{1}{8}z^{-1}+z^{-2})(1+1\frac{1}{2}z^{-1}+z^{-2})}{(1-\frac{5}{8}z^{-1}+\frac{5}{8}z^{-2})}$$

showing that useful designs can be obtained with quite coarse coefficient quantization. Figure 5 is the response of a filter with the pulse transfer function

$$G(z) = \frac{(1+\frac{5}{16}z^{-1}+z^{-2})(1-\frac{3}{32}z^{-1}+z^{-2})\times \\ \times(1-\frac{13}{16}z^{-1}+z^{-2})(1-1\frac{3}{4}z^{-1}+z^{-2})}{(1+\frac{3}{4}z^{-1}+\frac{7}{8}z^{-2})(1+\frac{3}{4}z^{-1}+\frac{17}{32}z^{-2})}$$

showing how far more stringent requirements can be met.

5.2 *The Design of Band-pass Filters*

Band-pass filters are inherently more difficult to design than low- or high-pass filters because a non-recursive section used in one stop-band will cause gain relative to the pass-band in the other stop-band. However, the above method can be applied to the band-pass case provided that this fact is taken into account in positioning the non-recursive section templates. An example of the response of a band-pass filter designed in this way is given in Fig. 6. The pulse transfer function is

$$G(z) = G_1(z)/G_2(z)$$

where

$$G_1(z) = (1-1\tfrac{15}{16}z^{-1}+z^{-2})(1-1\tfrac{7}{16}z^{-1}+z^{-2})\times \\ \times(1-1\tfrac{11}{16}z^{-1}+z^{-2})\times \\ \times(1+\tfrac{7}{16}z^{-1}+z^{-2})(1+1\tfrac{15}{16}z^{-1}+z^{-2})\times \\ \times(1+\tfrac{7}{8}z^{-1}+z^{-2})(1+1\tfrac{1}{2}z^{-1}+z^{-2})$$

and

$$G_2(z) = (1-z^{-1}+\tfrac{13}{16}z^{-2})(1+\tfrac{1}{8}z^{-1}+\tfrac{7}{8}z^{-2}).$$

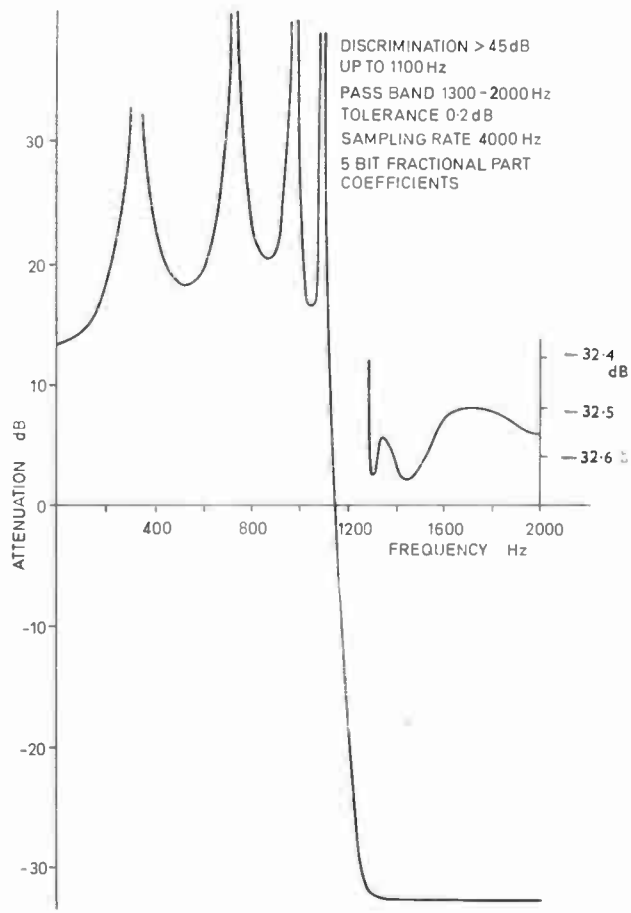


Fig. 5. Frequency response of a high-pass filter.

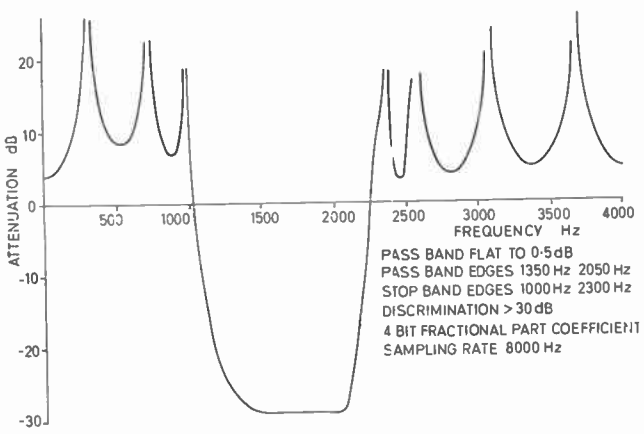


Fig. 6. A band-pass characteristic.

6. Conclusion

A template method for the design of digital filters has been described and design examples are given. The method offers the advantage of being able to take coefficient quantization into account. There is great scope for developing the use of the template method to design different types of filter.

7. Acknowledgments

The authors would like to thank the Director of the Post Office Research Department for permission to publish this work.

8. References

1. Jackson, L. B., Kaiser, J. F. and McDonald, H. S., 'An approach to the implementation of digital filters', *Trans. Inst. Elect. Electronics Engrs on Audio and Electroacoustics*, AU-16, pp. 413–21, September 1968.
2. Kaiser, J. F., 'Some practical considerations in the realization of linear digital filters', Proc. 3rd Allerton Conf. on Circuit and System Theory, pp. 621–33, 1965.
3. Scowen, F., 'An Introduction to the Theory and Design of Electric Wave Filters', p. 72 (Chapman and Hall, London, 1950).

9. Appendix

The Amplitude Gain/Frequency Characteristic of a Non-Recursive Quadratic Section

The frequency response of a network is defined by the relationship

$$G(j\omega) = Y(j\omega)/X(j\omega)$$

where $X(j\omega)$ is the spectrum of the input signal and $Y(j\omega)$ of the output signal.

In the case of digital filters $G(j\omega)$ is obtained by putting $z = e^{j\omega T}$ in the pulse transfer function $G(z)$.

Thus for a non-recursive quadratic section

$$G(z) = 1 + Az^{-1} + Bz^{-2}$$

$$G(j\omega) = 1 + Ae^{-j\theta} + Be^{-j2\theta}$$

where

$$\theta = \omega T$$

$$G(j\omega) = (1 + A\cos\theta + B\cos 2\theta) - j(A\sin\theta + B\sin 2\theta).$$

The amplitude gain/frequency characteristic $|G(j\omega)|$ is given by

$$|G(j\omega)|^2 = (1 + A\cos\theta + B\cos 2\theta)^2 + (A\sin\theta + B\sin 2\theta)^2$$
$$= A^2 + (1-B)^2 + 2A(1+B)\cos\theta + 4B\cos^2\theta.$$

Therefore,

$$|G(j\omega)| = [A^2 + (1-B)^2 + 2A(1+B)\cos\theta + 4B\cos^2\theta]^{\frac{1}{2}}.$$

Manuscript received by the Institution on 24th January 1972. (Paper No. 1481/CC 151).

The Authors

Mr. E. R. Broad read mathematics and physics at Oxford University and then joined Murphy Radio at Welwyn Garden City in 1934. In 1936 he entered the Post Office Engineering Department and was posted to the Radio Branch at Dollis Hill, where he worked on the design of filters, including crystal filters, for the London–Birmingham Coaxial Cable Terminals. During the war years, he was at an outstation at Banbury and he returned to Dollis Hill in 1948. There he led a group engaged on network design until ill-health caused early retirement in 1972.

Mr. P. F. Adams joined the Post Office in 1966 as a student apprentice. In 1967 he went to Southampton University where he read electronic engineering, graduating with an honours degree in 1970. He then joined the Post Office Research Department at Dollis Hill, working on the design and implementation of digital filters.

U.D.C. 621.396.621: 621.376.23

Synthesis of an Optimal Receiver Structure for Amplitude Modulated Pseudo-noise Signals

V. A. CHERDYNTSEV, B.Sc., C.Sc.,*
M. A. KHAN, B.Sc. (Graduate)†
and
A. R. MEMON, B.Sc., M.Sc., Ph.D.†

SUMMARY

An optimal discrete-time receiver structure is derived for demodulation of an amplitude modulated pseudo-noise carrier in the presence of white Gaussian noise. The receiver delivers maximum *a posteriori* estimates for the modulating information process and for the delay of the carrier.

* *Radiotechnical Institute, Minsk, 69, U.S.S.R.; on study leave at The City University, London.*

† *Department of Electrical and Electronic Engineering, The City University, St. John Street, London, EC1V 4PB.*

1. Introduction

In a number of telemetry applications in satellite and space communications, it is desirable to use a wideband sub-carrier (such as the well known pseudo-noise carriers based on pseudo-random binary sequences) instead of a sinusoidal carrier. In this short contribution, the synthesis of a receiver structure for such signals is described. The signal is assumed to comprise a binary pseudo-noise sub-carrier, amplitude modulated by a binary signal (the latter being the information-bearing message sequence), embedded in white Gaussian noise of known spectral density. This represents the video signal obtained by envelope detection of a received v.h.f. signal conveying the pseudo-noise sub-carrier. Such signals are used principally in space communications, where the assumption that the noise is Gaussian is generally valid.[1, 2]

The receiver also provides an estimate for the unknown delay in the pseudo-noise carrier (which may arise from the unknown and possibly changing distance between the transmitter and receiver).

By operating on a sampled version $r(kT)$ of the received signal $r(t)$, the receiver can be implemented by digital components, which may be expected to overcome difficulties which are encountered in the implementation of optimal non-linear filters by analogue techniques, and which have prevented the practical use of such non-linear filters. The sampled, received signal $r(kT)$ may be expressed as follows:

$$r_k = (a+h_k)g(t_k-\tau_k)+n_k \qquad \text{.....(1)}$$
$$= x_k+n_k$$

where

$$r_k \equiv r(kT), \qquad k = 0, 1, 2, \ldots, N, \qquad \text{.....(1a)}$$

and

$$t_k \equiv kT.$$

In the above equations T is the sampling interval and a is a known constant greater than unity (equal to the unmodulated sub-carrier amplitude) g and h are, respectively, the pseudo-noise carrier and the information-bearing message sequence, each having the two levels $(+1, -1)$, τ is the unknown delay, which may vary slowly, in a random manner, and is modelled by a continuous Markov process and n_k is sampled white noise of known spectral density $N_0/2$.

2. Estimation of the Message Signal

The derivation of the receiver structure when x_k (the noise-free modulation signal) is a two-level process and a is a known constant, is straight forward. Here a more general case in which $a > 1$ and x_k assumes four levels, namely, $\pm(a+1)$, $\pm(a-1)$, is considered. Expressions for the *a priori* and *a posteriori* probabilities of the quantities of interest are needed. P^+, P^-, will be used to denote respectively, the *a priori* probabilities of $+1$, -1, levels, and the *a posteriori* probabilities will be denoted by $\tilde{W}^+$, $\tilde{W}^-$ if non-normalized and by $\tilde{P}^+$, $\tilde{P}^-$ if normalized. Subscripts will be used to denote the particular process being referred to (e.g. g, h, or x).

The *a priori* probabilities of h and g are given by:[3, 4]

$$\left.\begin{aligned}\Delta P_h^+ &= -\Delta P_h^- = -\eta_h T P_h^+ + \zeta_h T P_h^- \\ \Delta P_g^+ &= -\Delta P_g^- = -\eta_g T P_g^+ + \zeta_g T P_g^-\end{aligned}\right\} \quad \text{.....(2)}$$

where ΔP denotes the difference between the probabilities at the $(k+1)$th and the kth sample, e.g.

$$\Delta P = P(kT+T) - P(kT). \quad \text{.....(3)}$$

To simplify subsequent algebra, it is convenient to define variables A and B as follows

$$\left.\begin{aligned}A &= \eta_h P_h^+ + \zeta_h P_h^- \\ B &= \eta_g P_g^+ + \zeta_g P_g^-.\end{aligned}\right\} \quad \text{.....(4)}$$

Thus, in the above equation P_h^+, P_h^- denotes respectively the probability of occurrence of a $+1$ and -1 level for h, and η_h, ζ_h are constants giving the probability of a change of level per unit time. $P_g^+, P_g^-, \eta_g, \zeta_g$ have corresponding meanings in relation to g.

Since x is a four-level process, four *a priori* probabilities are involved, and are given by

$$\begin{aligned}P_{x_1} &= P_h^+ P_g^+ \\ P_{x_2} &= P_h^+ P_g^- \\ P_{x_3} &= P_h^- P_g^+ \\ P_{x_4} &= P_h^- P_g^-. \end{aligned} \quad \text{.....(5)}$$

Using eqns. (2) and (5), and *a priori* probability for each level in discrete form is

$$\begin{aligned}\Delta P_{x_1} &= T(AP_g^+ + BP_h^+) \\ \Delta P_{x_2} &= T(AP_g^- - BP_h^-) \\ \Delta P_{x_3} &= T(-AP_g^+ + BP_h^-) \\ \Delta P_{x_4} &= T(-AP_g^- - BP_h^-). \end{aligned} \quad \text{.....(6)}$$

Consider now the non-normalized *a posteriori* probabilities for each level, which are given by[3, 4]

$$\Delta \tilde{W}_{x_i} = [\Delta P_{x_i}]_{\tilde{W}} - \frac{T}{2N_0}(x_i^2 - 2x_i r)\tilde{W}_{x_i}; \qquad i = 1, 2, 3, 4 \quad \text{.....(7)}$$

where to evaluate $[\Delta P_{x_i}]_{\tilde{W}}$, replace P by W in r.h.s. of eqn. (6). The normalized *a posteriori* probabilities are defined by

$$\tilde{P}_{x_i} = \frac{\tilde{W}_{x_i}}{\sum_{i=1}^{4} \tilde{W}_{x_1}} \qquad i = 1, 2, 3, 4 \quad \text{.....(8)}$$

therefore using eqns. (7) and (8) we obtain

$$\Delta \tilde{P}_{x_i} = [\Delta P_{x_i}]_{\tilde{P}} - \frac{T}{2N_0}\tilde{P}_{x_i}(x_i^2 - 2x_i r) + \sum_{\varepsilon=1}^{4} P_{x_\varepsilon}(x_\varepsilon^2 - 2x_\varepsilon r)$$
$$i = 1, 2, 3, 4 \quad \text{.....(9)}$$

where $[\Delta P_{x_i}]_{\tilde{P}}$ means replace P by $\tilde{P}$ in r.h.s. of eqns. (6). From eqn. (5) it can be seen that

$$\begin{aligned}P_h^+ &= P_{x_1} + P_{x_2} \\ P_h^- &= P_{x_3} + P_{x_4}.\end{aligned}$$

Further, letting

$$Z_h = P_h^+ - P_h^-, \qquad Z_g = P_g^+ - P_g^-$$

and using the fact that eqn. (5) implies an ordering of level notation such that the equation for the optimum non-linear filter from eqn. (9) becomes

$$\Delta Z_h = T(\zeta_h - \eta_h) - T(\zeta_h + \eta_h)Z_h + \frac{T}{4N_0}[(x_3^2 - x_1^2) - 2rZ_g(x_3 - x_1)](1 - Z_h^2). \quad \text{.....(10)}$$

Here, Z_g may be replaced by the estimate $g(t_k - \tau_k)$, which is a locally-generated version of the pseudo-noise sub-carrier produced by a feedback shift register and synchronized by a delay-lock-loop (see Section 3 and Fig. 1).

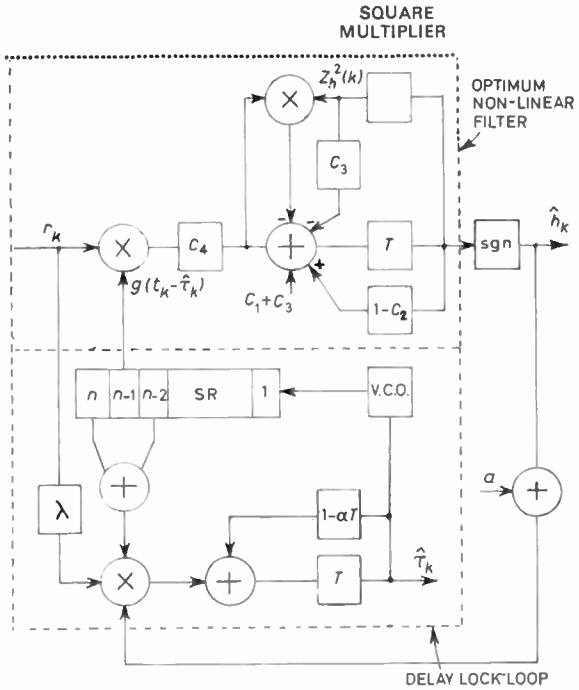


Fig. 1. Signal-flow diagram based on eqns. (12), (17).
T = one-sample delay.
VCO = voltage controlled oscillator.
SR = shift register.

The estimate h can be obtained simply from

$$\hat{h} = \text{sgn}\,(Z_h). \quad \text{.....(11)}$$

To facilitate the realization of an optimum non-linear filter, eqn. (10) can be rewritten in the following form:

$$\Delta Z_h = C_1 + C_2 + C_3 Z_h - C_2 Z_h^2 + C_4 r g(t_k - \hat{\tau}_k)[1 - Z_h^2] \quad \text{.....(12)}$$

where

$$\begin{aligned}C_1 &= T(\zeta_h - \eta_h) \qquad C_2 = T(x_3^2 - x_1^2)/4N_0 \\ C_3 &= T(\zeta_h + \eta_h) \qquad C_4 = T(x_3 - x_1)/4N_0.\end{aligned}$$

A part of Fig. 1 shows the signal flow diagram of the optimum non-linear filter, derived directly from eqn. (12).

3. Estimation of the Delay

An estimator for the delay τ is required and may be obtained as follows. Assume that τ is characterized by a first-order process:

$$\Delta\tau = -T\alpha\tau + Tn^\tau \quad \text{.....(13)}$$

where n^τ denotes white Gaussian noise of known spectral density and α is a constant, which represents the spectral width of τ. Using a Gaussian approximation for *a*

posteriori probability density $\Delta\tau$, the equation for the estimate $\hat{\tau}$ becomes[4]

$$\Delta\hat{\tau} = T\alpha\hat{\tau} + T\sigma_\tau^2 F_\tau \qquad \text{.....(14)}$$

where σ_τ^2 is the variance of the *a posteriori* estimate, and F_τ is given by

$$F_\tau = \frac{2}{N_0}\, r[a+\hat{h}]\,\frac{\partial g}{\partial \tau}(t_k - \hat{\tau}_k). \qquad \text{.....(15)}$$

For practical realization the partial derivative may be approximated as follows:

$$\frac{\partial g}{\partial \tau}(t_k - \hat{\tau}_k) \simeq [g(t_k - \hat{\tau}_k + T_1) - g(t_k - \hat{\tau}_k - T_1)]/2T_1 \qquad \text{.....(16)}$$

where T_1 = clock period of sequence g, and therefore from eqns. (14), (15) and (16), the equation for the estimator for τ becomes:

$$\Delta\hat{\tau} = -T\alpha\hat{\tau} + \lambda r[a+\hat{h}][g(t_k - \hat{\tau}_k + T_1) - g(t_k - \hat{\tau}_k - T_1)] \qquad \text{.....(17)}$$

where

$$\lambda = T \cdot \frac{\sigma_\tau^2}{N_0 T_1}.$$

In practice the period of the sequence may be decided in conjunction with the sub-carrier frequency from ranging requirements, i.e. if the estimate is to be used to determine the distance from transmitter to receiver the period must be long enough to give an unambiguous result. Otherwise, a shorter sequence would be used to facilitate initial acquisition. This completes the synthesis of the estimator and its structure corresponds closely to the conventional delay-lock loop.[2] It follows from eqns. (12) and (17) that the complete receiver structure will be as shown in Fig. 1.

4. Discussion

The receiver structure has been derived for the case $a > 1$, which corresponds to binary amplitude modulation of the sequence with a modulation depth of less than 100% (Fig. 2(a)). The same structure is applicable to 100% amplitude modulation if $a = 1$ (Fig. 2(b)) and to sequence-inversion modulation[1] (Fig. 2(c)) if $a = 0$.

Optimal receiver structures containing non-linear filters operating on continuous signals have been described previously, but are difficult to implement. The Fig. 1 structure operates as a sampled-version of the signal, and the implementation by conventional digital techniques should be straightforward.

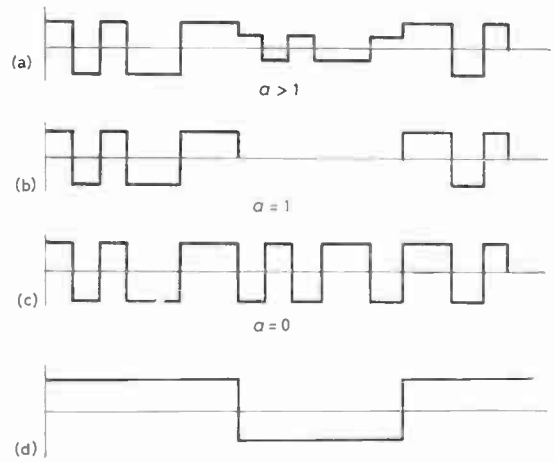


Fig. 2.

(a) Sequence with amplitude modulation depth less than 100% ($a > 1$).
(b) Sequence with amplitude modulation depth exactly 100% ($a = 1$).
(c) Sequence inversion modulation ($a = 0$).
(d) Modulation data-sequence.

5. Acknowledgment

One of the authors (M. A. K.) wishes to acknowledge the financial support of the Science Research Council (U. K.) in carrying out the work described.

6. References

1. Ward, R. B., 'Digital communications on a pseudo noise tracking link using sequence inversion modulation', *I.E.E.E. Trans. on Communication Technology*, **COM-75**, pp. 69–78, February 1967.
2. Golomb, S. W., 'Digital Communications' (Prentice Hill, Englewood Cliffs, New Jersey 1964).
3. Sosulin, Yu. G., 'Optimum reception of random pulse signals in noise', *Radiotekhnika i Electronika (USSR)*, **12**, p. 797, May 1967 (English translation in *Radio Engineering & Electronic Physics*, **12**, No. 5, p. 745, May 1967).
4. Cherdyntsev, V. A. and Davies, A. C., 'Synthesis of the optimal receiver structure for binary phase-modulated signals', *Electronics Letters*, **7**, pp. 739–41, 16th December 1971.

Manuscript first received by the Institution on 27th March 1972 and in revised form on 5th July 1972. (Short Contribution No. 159/Com. 57.)

U.D.C. 621.394.142: 621.395.34

An Experimental Adaptively Equalized Modem for Data Transmission over the Switched Telephone Network

R. J. WESTCOTT,
C.Eng., M.I.E.E.*

Presented at the IERE Conference on Digital Processing of Signals in Communications held in Loughborough from 11th to 13th April 1972.

SUMMARY

The relative merits of the zero-forcing, mean-square error and quantized-feedback algorithms are compared for data transmission over the switched telephone network, bearing in mind the nature of the transmission impairments met in this medium. The design and performance of an experimental 4-level vestigial-sideband amplitude-modulated modem operating at a data signalling rate of 4800 bit/s with a hybrid 37-tap equalizer using the mean-square error algorithm for 6 leading taps and quantized feedback for 30 trailing taps is discussed.

* *Post Office Research Department, Dollis Hill, London NW2 7DT.*

1. Introduction

The present rate of increase of computer and data processing services in the United Kingdom is such that to satisfy the demand for data communications the British Post Office will need to cater for a growth rate that effectively doubles the number of data terminals each year. The ubiquitous public switched telephone network which provides communication between, practically, any two points in the United Kingdom will, for many years to come, play an important part in meeting the demand. The existing Datel services have shown that the telephone voice channel is a medium capable of providing economic data communication facilities with adequate performance for the majority of customers. However, due to the transmission impairments that exist in the switched telephone network and the desire, initially, to avoid exact prescription of data signalling rates, the service offered at present is restricted to an upper limit of 1200 bit/s.

The public switched telephone network has, of course, been designed to meet the requirements of speech transmission, where the optimization of the network transmission characteristics has been mainly concerned with achieving a good noise performance and adequate reproduction of the power spectrum of speech; waveform distortion caused by amplitude-frequency and phase non-linearities or listener echo of delay less than 30 ms not being of prime significance. As a result of this, it is waveform distortion that limits the data signalling rate in the switched telephone network rather than the noise which, itself, is sufficiently good to enable much higher data rates than those currently offered. The effect of this waveform distortion is to spread out in time the response of a data pulse such that it overlaps adjacent pulses in a digital wavetrain. This is called intersymbol interference and it reduces the margin against noise or, in severe cases, causes systematic decision errors.

In this paper various forms of waveform correctors for reducing intersymbol interference are described and their relative merits discussed, bearing in mind the nature of the transmission impairments met in the switched telephone network.

This type of equalizer optimizes the performance in the time domain, minimizing the intersymbol interference at the sampling instants. As the magnitude and nature of the distortion is particular to each connexion the equalization process has to occur after each call has been set up. Also, in practice, the telephone channel cannot be regarded as completely time invariant. In view of this, it is desirable for the equalization to be a continuous process, where the waveform corrector continuously adapts itself by deriving the information required for adjusting the equalizer direct from the incoming data signal. In this manner the equalizer optimizes the performance during the whole period of data transmission.

Various algorithms, i.e. adaptive strategies, and their relative merits are discussed in this paper. A particular implementation of an adaptively equalized modem using vestigial-sideband amplitude modulation with a mean-square-error algorithm for pre-pulse correction and

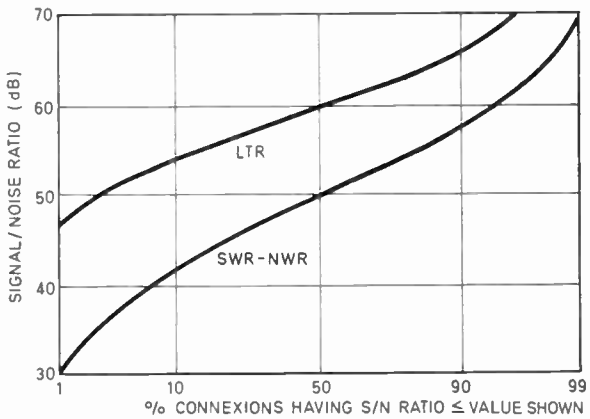


Fig. 1. Cumulative distribution of signal/noise ratio. (LTR—London Telephone Region, SWR—South-West Region, NWR—North-West Region).

quantized feedback for post-pulse correction is described, and its measured performance is discussed.

2. Transmission Impairments in the Switched Telephone Network

2.1. *Noise*

The cumulative distribution of measured overall signal/noise ratio (subscriber-to-subscriber) as a function of percentage of connexions for general background noise[1] is shown in Fig. 1 for the London Telephone Region (LTR) and for trunk circuits between the South West and North West Regions (SWR–NWR). The median value of signal/noise ratio for trunks and the LTR is 50 dB and 60 dB respectively, with 99% of the trunk connexions, which is the more severe case, having a signal/noise ratio better than 30 dB.

Assuming that the bandwidth in the telephone channel available for the data signal is about 2 kHz, then information theory gives an upper bound on capacity of about 20 000 bit/s for a signal/noise ratio of 30 dB, which is significantly higher than the 600/1200 bit/s data signalling rate currently provided by the Datel service. It is obvious from this that it is not background noise that is limiting the data signalling rate but rather the intersymbol interference caused by phase non-linearities, amplitude-frequency distortion and echo. The application of adaptive equalization techniques to reduce this intersymbol interference will enable the use of multi-level transmission systems having data signalling rates that compare much more favourably with the theoretical upper bound.

The mean error rate for 2-level data transmission systems using the switched telephone network is of the order of 1 in 10^4 and is caused by impulsive noise primarily due to the electro-mechanical switches in the telephone exchange. Due to the nature of the impulsive noise a change of about 10 dB in signal/impulsive-noise ratio is required to give an order change in error rate. In view of this, and noting that we are comparing an equalized multi-level system with a distorted 2-level system, it is thought that the mean error rate of an adaptively equalized 4-level system will not be significantly different from that of the present Datel service. For a higher number of levels, however, the error rate will increase and for those subscribers who require mean error rates better than 1 in 10^4 error correction systems will be required.

2.2. *Amplitude and Phase Distortion*

For pulses to be transmitted at a rate $1/T$ without intersymbol interference the voltage spectrum of the received signal should have a low-pass amplitude characteristic with an odd-order symmetrical roll-off about the Nyquist frequency, $F = 1/2T$, and a phase characteristic that is linear with frequency. Departure from this ideal state, due to non-uniform amplitude-frequency characteristics and phase non-linearities in the telephone channel, causes significant intersymbol interference and is one of the main reasons for the limited data signalling rates of the Datel service. The effect can clearly be seen in Fig. 2, where the measured response of a single pulse is shown for a vestigial-sideband amplitude-modulated system with a modulation rate of 2400 bauds after traversing two carrier systems and 100 miles of loaded cable.

The unit pulse of period T has been dispersed in time for a duration of about eight elements, giving significant values of intersymbol interference for four sampling instants, h_{-4} to h_{-1}, preceding the main sample h_0, and about four succeeding samples h_1 to h_4. This is typical of the form of waveform distortion caused by non-linear phase and amplitude characteristics. The magnitude of the distortion depicted in Fig. 2 is such that binary transmission would be marginal, the binary eye† being impaired by some 14 dB, and multi-level operation would be impossible without the aid of adaptive equalization techniques to reduce the intersymbol interference.

2.3. *Listener Echo*

An important form of distortion met in the public switched telephone network that is not shown in Fig. 2 is 'listener echo' caused by reflexions due to inadequate

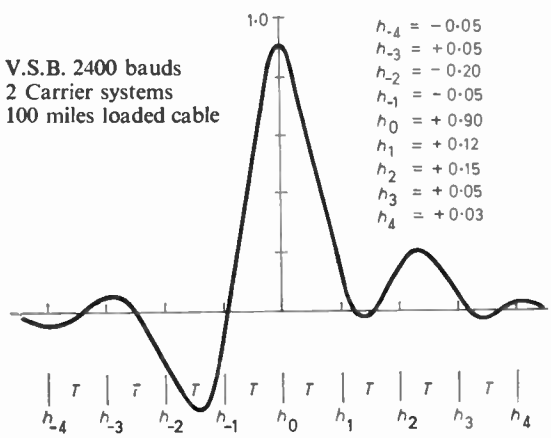


Fig. 2. Measured pulse response.

† The 'eye pattern' is a convenient way of displaying on an oscilloscope the aggregate intersymbol interference over an element period when transmitting a wavetrain. For an explanation of 'eye pattern' see pages 119 and 287 of reference 10.

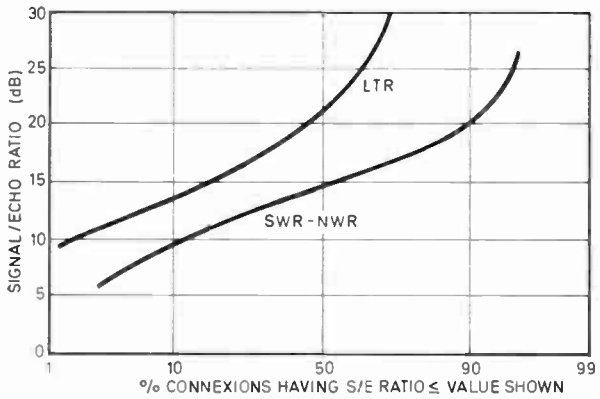


Fig. 3. Cumulative distribution of signal/echo ratio.

balance-return loss of 2-wire/4-wire hybrids and mismatches in loaded junction cables etc. The cumulative distribution of the worst signal/echo ratio measured[1] in the frequency band 900 to 2400 Hz for the switched telephone network as a function of percentage of connexions is shown in Fig. 3 for the London Telephone Region and trunks. This form of impairment can be very severe, the median value of signal/echo ratio for trunk calls being about 14·5 dB, with 10% of the connexions having a signal/echo ratio of less than 10 dB. It is thought that realignment of certain of the trunk circuits will improve the median value by 3 to 4 dB and the 5% point by about 6 dB. However, noting that a signal/echo ratio of 10 dB is sufficient to cause systematic error in a 4-level system, it can be seen that listener echo is one of the main degrading factors limiting the data signalling rate in the switched telephone network.

Echoes having delays less than 30 ms are not usually of great importance for speech transmission[2] but for data we are very much concerned with its effect at all the sampling instants. Measurements of echo delay on trunk circuits indicate that 1% of connexions will have delays greater than 12 ms. In fact, to cater for all possible connexions in the United Kingdom, the adaptive equalizer must be capable of equalizing pulses which have been dispersed in time following the main sample by up to about 15 ms. This represents many bit intervals, e.g. the adaptive equalizer will be required to minimize intersymbol interference over a range of 27 trailing samples for a modulation rate of 1800 bauds and 36 trailing samples for a modulation rate of 2400 bauds. This is an important point to consider when comparing the various adaptive algorithms both in terms of technical performance and economic viability.

3. Network Topology for Time Domain Equalization

3.1. *The Tapped Delay Line Transversal Filter*

The tapped delay line transversal filter as a time domain equalizer has a history dating back some 40 years. As can be seen from Fig. 4 it is, conceptually, a simple device which is ideally suited for the purpose of adaptive equalization. It comprises a linear delay line with non-recursive taps uniformly spaced by T seconds, the outputs of which are summed after appropriate weighting by the variable tap gains, $c_j, -n < j < p$, to give the equalized output signal Y. It follows that the output sequence, Y, is given by the convolution of the impaired input pulse sequence, X, with the tap gain sequence, C, i.e.

$$y_k = \sum_j c_j x_{k-j} \qquad -n < j < p \qquad \text{.....(1)}$$

or, alternatively, using matrix notation $y_k = C'X_k$ where C and X are regarded as column vectors and C' is the transpose. By adaptively controlling the tap gains, c_j, the output signal can be equalized to meet some suitable performance criterion.

The linear delay line transversal filter can be used for equalizing analogue or digital signals. For isochronous data transmission the time delay T is made equal to the reciprocal of the modulation rate. It can be seen that the tap gains c_j, which can be positive or negative in sign, act directly in the time domain to minimize the intersymbol interference at the sampling times h_j in the pulse response. Ideally, to provide the perfect inverse filter, an infinite number of taps are required. The practical necessity of truncating the number of taps results in a residual intersymbol interference, the magnitude and nature of which depends on the initial waveform distortion and the algorithm setting the value of the tap gains. The number of taps required must be sufficient to provide a range of time delay wider than the range of dispersion of the impaired signal X in order to ensure that the residual convolved values of intersymbol interference are reduced to a satisfactory level.

It can be shown that a more general network which includes recursive taps will equalize to a given residual intersymbol interference with a shorter delay line. However, to ensure stability there have to be restrictions on the magnitude of the recursive taps. This is a considerable complication to include in the algorithm for setting the tap gains and, as a result of this, linear recursive taps are not favoured for adaptive equalization. By using quantized feedback, however, the recursive form of network discussed in the next section becomes a very attractive proposition.

It should be noted that the transversal filter increases the additive noise, each tap making its contribution to the total output noise level. Assuming the noise samples at each tap are uncorrelated then the noise contributions will add on a power basis such that the additive noise is enhanced by the factor

$$\sum c_j^2 \qquad -n < j < p. \qquad \text{.....(2)}$$

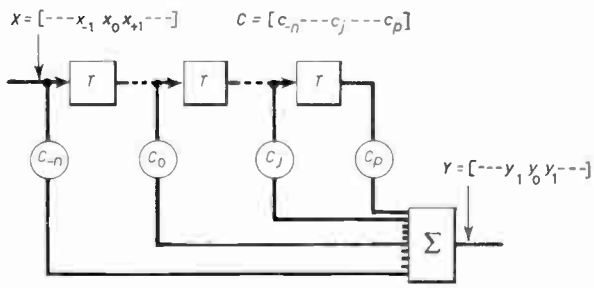


Fig. 4. Linear non-recursive transversal network.

3.2. *Quantized-feedback Correction*

A recursive form of network using quantized feedback that is suitable for digital signals only is shown in Fig. 5. Stability is ensured by the action of the quantizer in the feedback loop, the signal being quantized to defined discrete levels determined by the number of levels in the digitally-modulated signal. The quantizer is, of course, the decision device which recovers the original message sequence M. This is then passed to the delay line which, as we are now only concerned with discrete levels, can be in the form of a simple shift register. As for the linear transversal filter, the summation of the weighted tap outputs is subtracted from the impaired signal X to give the corrected signal Y. However, as we are now subtracting ideal undistorted signals (the linear transversal filter subtracts delayed versions of the impaired signal) convolution does not occur and the additive noise is not enhanced. Due to the absence of convolution a much shorter delay line can be used to achieve a satisfactory level of residual intersymbol interferences.

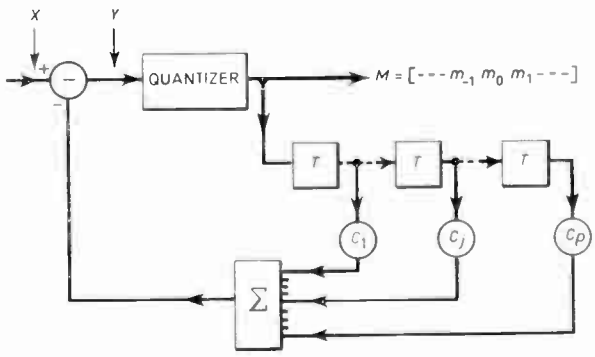


Fig. 5. Quantized-feedback correction.

The disadvantage of this form of network topology is that error extension can occur, i.e. dependent on the magnitude of the tap gains, c_j, further errors can be caused as a decision error propagates along the delay line. Also, there is the possibility of lock-out if decision errors are excessive and persist for some time, i.e. the tap gains could settle to values that generate a cyclic repetitive pattern in the delay line. This is a possibility during the initial phase of equalizing highly distorting lines where the decision error rate can be very high. However, as discussed in the measurements section, experience has shown that this is not a problem for 4-level operation in the public switched telephone network.

It is obvious that the quantized-feedback waveform corrector is only suitable for correcting intersymbol interference following the main pulse sample. As the majority of pulse dispersion in the public switched network follows the main pulse sample and echoes having delay times up to 15 ms need to be considered, the quantized-feedback corrector is, both in terms of technical performance and economic viability, an attractive proposition.

4. Algorithms for Setting Tap Gains

Over the past few years several algorithms, or strategies, have been developed for the adjustment of the tap gains of a transversal filter. Important contributions to this field have been made by Lucky[3,4] who describes an algorithm that forces zeros over a truncated range of the channel impulse response. Another very important algorithm that minimizes the mean-square-error of the intersymbol interference is described by Niessen[5,6] for data transmission, where the adjustment of the tap gains is made a continuous process throughout transmission by using a decision-directed error signal. A similar algorithm is described by Lucky and Rudin[7] for analogue signals using an ideal reference training signal. The pertinent features of these algorithms and variants of them are discussed in this section.

4.1. *The Zero-forcing Algorithm*

Lucky defines a criterion of performance for setting the tap gains that minimizes a distortion factor,

$$D = \sum_{j \neq 0} |h_j|,$$

where h_j is the equalized sampled pulse response. This is equivalent to maximizing the eye opening. It is shown that if the initial distortion is such that the binary eye is open, then this criterion is met by adjusting tap gain c_j to force a zero in the equalized channel impulse response at h_j, $j \neq 0$, the main tap c_0 being adjusted to give $h_0 = 1$. Thus, an equalizer with n leading taps and p trailing taps will force $n+p$ zeros.

It can be shown that an estimate $\hat{h}_j$ of the equalized channel impulse response is given by the correlation between the message sequence and an error sequence, i.e.

$$\hat{h}_j = \overline{m_{i-j} e_i} \qquad \text{.....(3)}$$

where m_{i-j} is the message sample, $e_i = (y_i - m_i)$ is the error sample, and the product is averaged over N samples. It is assumed that the input data is uncorrelated.

Lucky has shown that the distortion function D is convex with one global minimum. This allows iterative procedures to be used for incrementing the tap gains to their optimum values; the zero-forcing algorithm adjusting the tap gain c_j in increments ∂c_j such that $\hat{h}_j$, $j \neq 0$, becomes zero. A particular implementation of this is shown in Fig. 6 where the tap increment is proportional and opposite in sign to the estimate $\hat{h}_j$, i.e.

$$\partial c_j = -K \overline{m_{-j-n} e_{-n}} \qquad \text{.....(4)}$$

where the error sequence E has been delayed by nT to enable correlation measurement for the n leading taps.

In Lucky's implementation of the zero-forcing algorithm the signum functions (polarity only) of m and e are used to determine ∂c_j, i.e.

$$\partial c_j = -\Delta \overline{\operatorname{sgn} m_{-j-n} \operatorname{sgn} e_{-n}}. \qquad \text{.....(5)}$$

Using this strategy the correlator multipliers become the easy to implement 'exclusive OR' function, whereas linear multipliers are required for the strategy defined by equation (4). However, the clipping required to derive the signum functions means some information has been lost, resulting in a possible penalty in the speed of convergence. Using the strategy of equation (4) the size of the incremental step is proportional to the estimate of intersymbol interference $\hat{h}_j$ and occurs at uniform

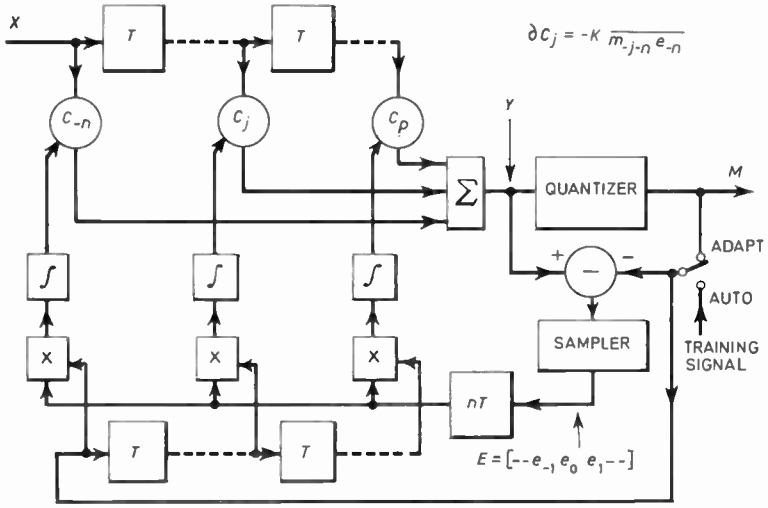


Fig. 6. Zero-forcing adaptive equalization.

intervals of NT seconds. This proportionality provides fast convergence when the error is large, but it produces small, ideally zero, residual tap jitter when optimum equalization has been achieved. In Lucky's implementation the incremental step is a constant Δ and, in order to optimize the speed of convergence and the residual tap jitter, an up-down counter is used to do the averaging. Incremental steps now occur at non-uniform intervals of time, the spacing being closer the larger the error signal.

A disadvantage of the zero-forcing algorithm is that, even with an ideal reference training signal, convergence to the optimum solution cannot be guaranteed if the distortion is sufficiently severe to close the binary eye. This is a condition that can occur in the public switched telephone network. Though, as discussed by Niessen,[6] convergence with a training signal can be ensured by preceding the zero-forcing equalizer with an adaptive matched filter, this increases the complexity of the equipment and also the magnitude of the intersymbol interference which the zero-forcing equalizer has to deal with. In the adaptive mode, when both the message sequence M and the error sequence E are decision directed, convergency will be affected by the error rate during the initial phase of equalization. Lucky has shown by computer simulation that the adaptive mode appears to be tolerant of error rates as high as 1 in 10.

4.2. *The Mean-square-error Algorithm*

The criterion of performance for the mean-square-error algorithm is defined as the minimization of the mean-square-error between the channel output y_k and the transmitted message m_k, i.e. minimization of

$$\overline{e_k^2} = \overline{(y_k - m_k)^2}. \qquad \text{.....(6)}$$

It can be shown that the mean-square-error is a convex function of the tap gains with one global minimum such that the optimum solution for each tap is given by

$$\partial \overline{e_k^2}/\partial c_j = 0.$$

Hence, an iterative procedure can be used to adjust the tap gains to the optimum solution by incrementing in steps proportional and opposite in sign to the gradient

$$\partial \overline{e_k^2}/\partial c_j.$$

From equations (1) and (6) it follows that

$$\frac{\partial \overline{e_k^2}}{\partial c_j} = 2\overline{e_k x_{k-j}} = 2\overline{e_0 x_{-j}}$$

and the tap gain increment is given by,

$$\partial c_j = -K\overline{e_0 x_{-j}}. \qquad \text{.....(7)}$$

Thus, the strategy for incrementing the taps involves the measurement of the cross correlation between the error signal e_0 and the tap signal x_{-j}. A simplified

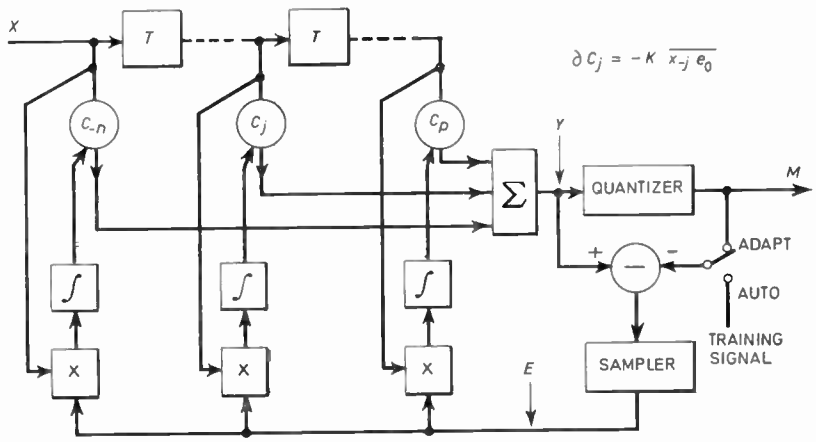


Fig. 7. Mean-square-error adaptive equalization.

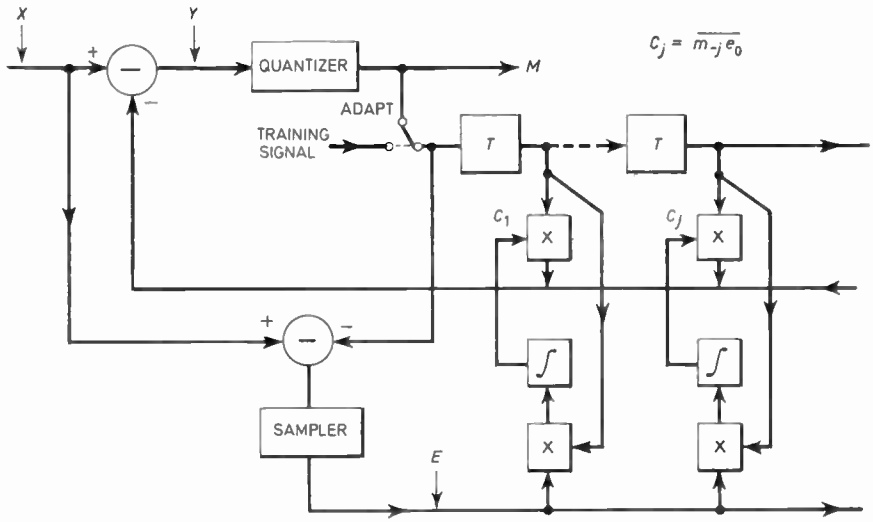


Fig. 8. Quantized-feedback coefficient.
(a) Renewing method.

functional diagram showing the basic principles of the algorithm is given in Fig. 7. Subject to an upper bound on the magnitude of the proportionality constant K, it can be shown that with an ideal reference training signal this algorithm will converge to the optimum solution for all distortions. In the adaptive mode, the error sequence E is decision directed and computer simulation indicates that the algorithm will still converge with error rates at least as high as 1 in 10.

The implementation depicted in Fig. 7 requires complex and relatively expensive correlator multipliers which are linear on both ports. However, it has been shown[8] that either or both e_0 and x_{-j} can be replaced by their respective signum functions without significantly affecting the residual equalized mean-square-error, though with a penalty on the speed of convergence.

4.3. *Quantized-Feedback Algorithms*

The basic strategy for adjusting the tap gains of a quantized-feedback equalizer is the same as that for the zero-forcing algorithm in as much as it is based on the measurement of correlation between the message sequence M and the error sequence E. This results, as previously discussed, in each tap c_j forcing a zero in the equalized pulse response. However, the significant difference between this and the linear transversal filter form of zero-forcing is that the correction signal is now a weighted sum of the message M rather than the impaired signal X. As convolution does not now occur, there is no induced intersymbol interference outside the range of the equalizer and the action of forcing zeros ensures that the optimum solution, i.e. the minimization of the distortion factor D, has been achieved. In fact, with an ideal reference training signal, this algorithm will converge to the optimum solution for all distortions including those cases where the binary eye is closed. For a given length equalizer, the residual intersymbol interference is, in general, less with quantized feedback and there is, of course, no enhancement of noise.

Two variants of the quantized-feedback algorithm are shown in Figs. 8(a) and (b). In the configuration shown in Fig. 8(a) the output of the correlator averaged over N samples gives the tap gain for forcing a zero and only a relatively inexpensive multiplier is required to give the correction signal instead of the more complicated incrementing memory type of multiplier. This follows from the fact that the mean value of the estimate $\overline{h_j}$ is equal to h_j the intersymbol interference of the unit

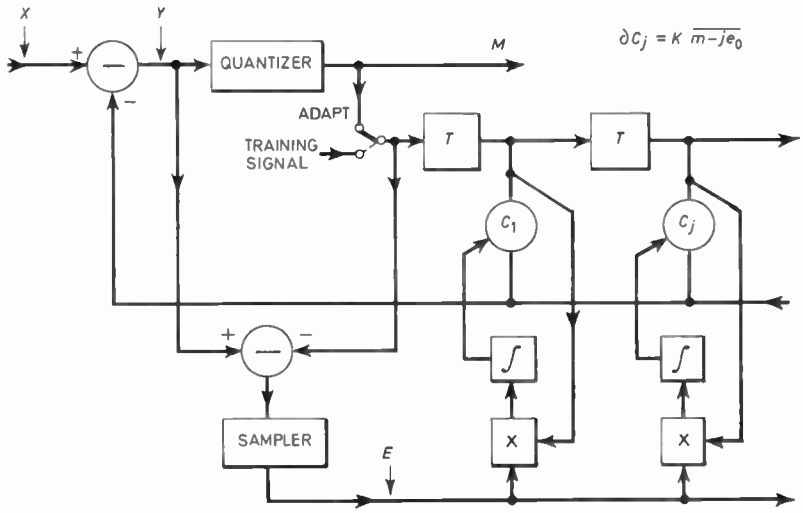


Fig. 8. Quantized-feedback coefficient.
(b) Incrementing method.

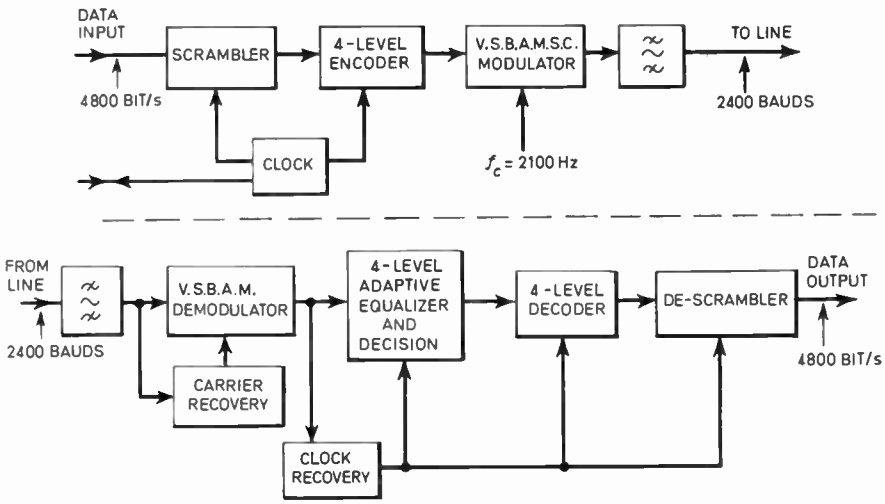


Fig. 9. An experimental adaptively equalized modem.

pulse at sampling time j. Thus, from equation (3),

$$\overline{h_j} = h_j = c_j = \overline{m_{-j}e_0}. \qquad \ldots\ldots(8)$$

As there are no incrementing memories, the error signal e_0 cannot go to zero and must be derived from signal X rather than Y, thus $e_0 = (x_0 - m_0)$ and not $(y_0 - m_0)$ as for the previous algorithms. A disadvantage of this open loop configuration is the increase in sensitivity to message statistics, necessitating a longer averaging period in the correlator than would be necessary with an incrementing memory system, resulting in a slower speed of convergence.

The configuration with an incrementing memory is shown in Fig. 8(b). In this case the tap gain is incremented to the optimum solution by the strategy,

$$\partial c_j = K\overline{m_{-j}e_0} \qquad \ldots\ldots(9)$$

where e_0 is now given by $(y_0 - m_0)$.

Computer simulation of the adaptive mode indicates that the quantized-feedback algorithm is as tolerant of decision error as are the zero-forcing and mean-square error algorithms, even though there is the possibility of error extension due to feedback of decision errors in the correction signal. In view of its good technical performance and the practical advantages of implementation using shift registers for delay lines, the use of quantized-feedback techniques to deal with the wide range of pulse dispersion following the main pulse sample is a very attractive proposition for the public switched telephone network.

5. An Experimental Adaptively Equalized Modem

5.1. *The Experimental Modem*

The simplified block schematic of the experimental adaptively equalized modem operating at a data signalling rate of 4800 bit/s is shown in Fig. 9. As it is desirable to have no restrictions on the format of the data input it is necessary to scramble the signal in order to minimize any message correlation in the 4-level encoded signal. The pseudo-random generators in the scrambler and descrambler that perform this function are self-synchronizing. The number of stages in the pseudo-random generator must be sufficient to ensure that the secondary auto-correlation peaks that occur in the 4-level encoded pseudo-random sequence are outside the range of the adaptive equalizer. For the measurements discussed in the next Section, a 9-stage, 511-bit, pseudo-random pattern was used which gives secondary auto-correlation peaks well outside the range of the 37-tap adaptive equalizer.

The 4-level encoded signal is transmitted on a carrier frequency of 2100 Hz at a modulation rate of 2400 bauds using vestigial-sideband amplitude modulation (v.s.b.a.m.) with suppressed carrier. A detailed investigation of the comparative performance of various modulation systems for the public switched telephone network, based on the measured range of transmission impairments discussed by Ridout and Rolfe,[1] indicates that v.s.b.a.m. is the optimum system. Quadrature amplitude modulation is comparable in performance but, as this is a two-channel system, it significantly complicates the adaptive equalizer and, in view of this, v.s.b.a.m. is preferred for an adaptively equalized modem. The spectrum of the data signal is from 600 to 2500 Hz and allows for a low-frequency return channel below 600 Hz.

The coherent carrier recovery system required for demodulation consists of an oscillator phase locked to a 2100 Hz pilot carrier, which is transmitted at a level of −6 dB relative to the data signal giving a penalty of 0·8 dB on noise performance. The phase-locked oscillator has been optimized in terms of its ability to track phase and frequency offset with tolerable residual phase jitter due to the data spectrum. The received pilot carrier is also used to control the overall gain of the system. The clock recovery method is the conventional zero crossing type with a digital up/down counter to average the early and late transitions.

This basic modem provides the test bed for investigating the various adaptive algorithms discussed in Section 4, allowing an assessment of their comparative performance for the whole range of transmission impairments met in the public switched telephone network. The particular 4-level adaptive equalizer to be discussed in Section 5.2 has a total of 37 taps, with 6 leading taps

using the mean-square error algorithm and 30 trailing taps using the quantized-feedback coefficient renewing method depicted in Fig. 8(a). Subsequently, this performance will be compared with that obtained using the mean-square error algorithm for all the taps and, also, the quantized-feedback coefficient incrementing method for the trailing taps. In addition to this the various adaptive strategies are being studied using computer simulation.

The majority of the experimental adaptively equalized modem has been implemented using digital circuitry. The scrambler/descrambler, 4-level encoder/decoder and a large number of the functions in the adaptive equalizer are implemented with t.t.l. integrated circuits. Linear integrated circuits are used for the active filters, the processes of modulation/demodulation and the correlator multipliers in the adaptive equalizer. The variable-gain function for the leading taps is achieved using field effect transistors as a variable resistor. Linear integrated circuits and field effect transistors are also used for the sample-and-hold type delay line for the leading taps.

Much of the circuitry could be implemented using medium scale integration. Probably, in the future, when the market is such that the use of large-scale integration becomes viable economically, the all-digital approach with re-circulating shift registers to perform the various mathematical functions will be the best method of implementation.

5.2. *Measured Performance*

Using the line transmission simulator discussed by Groves and Mackrill,[9] the performance of a modem in the presence of the various forms of distortion met in the public switched telephone network can be assessed. This enables, in combination with the statistical description of the transmission characteristics of the switched telephone network given by Ridout and Rolfe[1] which is based on a large sample (about 1000 test recordings) of measured lines, the performance of a modem to be defined in terms of the percentage of connexions that will give satisfactory performance. Various combinations of transmission impairments including amplitude-frequency distortion, group-delay distortion and listener echo covering the whole range met in the switched telephone network are set up on the line transmission simulator and the resulting eye impairment measured. Subsequent processing by computer enables the cumulative distribution of eye impairment as a function of percentage of connexions to be plotted. In addition to this, the performance of a modem in the presence of amplitude hits, phase hits, frequency offset, white Gaussian noise and impulsive noise can be assessed.

5.2.1. Cumulative distribution of eye impairment

The cumulative distribution of eye impairment as a function of percentage of connexions is shown in Fig. 10 for trunks and intra-regional calls in London, the South West and North West Regions. The curves have been drawn asymptotic to the 1·5 dB back-to-back eye

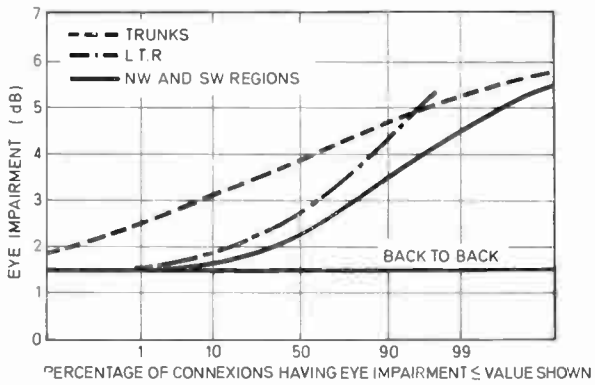


Fig. 10. Cumulative distribution of eye impairment.

impairment of the modem. The maximum eye impairment for any connexion is less than 6 dB giving a maximum degradation of about 4 dB relative to the back-to-back value, this probably being mainly due to the limited number of leading taps. However, this degree of impairment is quite acceptable in a practical system and it is doubtful that it is worthwhile having more than 6 leading taps. It should be noted that all the tests were performed with the adaptive equalizer in the 4-level mode and it successfully equalized 100% of the trunks and the connexions in the South West and North West Regions and 98% of the connexions in the London Telephone Region. The eight connexions out of the total 1000 line sample on which the equalizer failed to converge had distortion which was a combination of severe attenuation slope and an echo in the worst phase condition. It must be noted that the assessment is pessimistic in the sense that, for other echo phases, the equalizer would have converged to the optimum solution. These lines could be equalized by using a binary mode equalization procedure to reduce the initial systematic decision errors, returning to the multi-level configuration after the equalization period. Possibly, a preferable arrangement would be to stay in the 4-level mode and use a fixed compromise equalizer to reduce the maximum slope across the frequency band.

The total overall response time for a.g.c. recovery, clock recovery, carrier recovery and adaptive equalization is typically about 3 seconds for the range of distortion met in the switched telephone network. The median response time for the equalizer alone was < 500 ms.

5.2.2. Noise performance

The bit error rate as a function of the normalized signal-to-noise ratio (signal energy per bit to noise power per Hz) for white Gaussian noise is shown in Fig. 11. The theoretical performance for a 4-level vestigial-sideband amplitude modulated system is also shown for reference. It can be seen that the measured back-to-back performance is about 1·5 to 2·0 dB down relative to the theoretical. Of this, 0·8 dB is due to the pilot carrier, giving a modem noise performance within about 1 dB of theoretical, this being a result of the back-to-back eye impairment discussed in the previous section. To achieve this degree of performance the random tap jitter of the

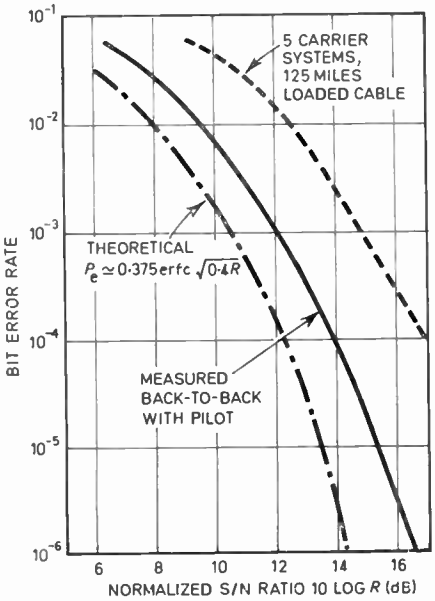


Fig. 11. Noise performance of modem.

37-tap equalizer has been minimized by using a time constant of 300 ms for the correlators in the quantized-feedback section of the equalizer. As the quantized-feedback coefficient incrementing method type of equalizer is less sensitive to message statistics, it is thought that the correlation time-constant can be reduced for this type of equalizer to give faster equalization times with tolerable tap jitter.

The effect of severe distortion on noise performance is shown in Fig. 11 for 125 miles of loaded cable and five carrier systems. The degradation on back-to-back performance is about 3 dB. The majority of this impairment is due to convoluted leading intersymbol interference outside the range of the equalizer, the noise enhancement effect of the leading taps being less than 0·5 dB. Noting from Fig. 1 that 99% of the connexions in the switched telephone network have a signal/noise ratio greater than 30 dB it can be seen that background Gaussian noise is not an important parameter for 4-level v.s.b. operation.

Tests with impulsive noise have shown that the adaptive equalizer convergence is not affected with mean bit error rates as high as 1 in 10, which is several orders higher than the 1 in 10^4 to 1 in 10^5 normally obtained on the switched network.

6. Conclusion

The first phase of the study of adaptive equalization strategies suitable for data transmission over the switched telephone network has shown that excellent performance can be obtained using quantized-feedback techniques. Based on a 1000-line sample of the switched network, it has been shown that a hybrid form of adaptive equalizer using the mean-square algorithm for the leading taps and quantized feedback for the trailing taps will converge in the 4-level mode for more than 99% of the connexions. Also, it has been shown that lock-out is not a problem and a large number of taps (37 in the experimental modem) can be used with minimal tap jitter giving a noise performance close to theoretical.

7. Acknowledgment

Acknowledgment is made to the Director of Research of the British Post Office for permission to publish this paper. Thanks are also due to the colleagues who participated in the work.

8. References

1. Ridout, P. N. and Rolfe, P., 'Transmission measurements of connexions in the switched telephone network', *Post Office Elect. Engrs. J.*, **63**, Part 2, p. 97, July 1970.
2. Richards, D. L., 'Theoretical study of the function of echo suppressors', *Teleteknik* (English Edition), 7, No. 2, p. 71, 1963.
3. Lucky, R. W., 'Automatic equalization for digital communication', *Bell Syst. Tech. J.*, **44**, p. 547, April 1965.
4. Lucky, R. W., 'Techniques for adaptive equalization of digital communication systems', *Bell Syst. Tech. J.*, **45**, p. 255, February 1966.
5. Niessen, C. W., 'Automatic channel equalization algorithm', *Proc. Inst. Elect. Electronics Engrs*, 55, No. 5, p. 698, May 1967. (Letters.)
6. Niessen, C. W. and Willim, D. K., 'Adaptive equalizer for pulse transmission', *I.E.E.E. Trans*, on *Communication Technology*, **COM-18**, No. 4, p. 377, August 1970.
7. Lucky, R. W. and Rudin, H. R., 'An automatic equalizer for general-purpose communication channels', *Bell Syst. Tech. J.*, **46**, p. 2179, November 1967.
8. Hirsch, D. and Wolf, W. J., 'A simple adaptive equalizer for efficient data transmission', *I.E.E.E. Trans. on Communication Technology*, **COM-18**, p. 5, February 1970.
9. Groves, K. and Mackrill, P., 'A line-transmission simulator for testing data transmission systems', *Post Office Elect. Engrs J.*, **63**, Part 2, p. 117, July 1970.
10. Bennett, W. R. and Davey, J. R., 'Data Transmission', Inter-University Electronics Series, (McGraw-Hill, New York, 1965).

Manuscript received by the Institution on 16th December 1971.
(Paper No. 482/Com. 58.)

U.D.C. 621.382.323

M.O.S.F.E.T. Temperature-Drift Performance Limitations

R. W. J. BARKER, M.Sc., C.Eng., M.I.E.E.*
and
B. L. HART, B.Sc., C.Eng., M.I.E.R.E.†

SUMMARY

An analysis, employing a minimum of restrictive assumptions, is presented that allows the temperature drift performance of m.o.s.f.e.t.s operated at a constant drain current in the pinch-off region to be assessed. The parabolic variation of gate-source voltage with temperature for devices working in the vicinity of the minimum drift bias point is related to device parameters, which are dependent on details of the manufacturing process. The detrimental effect of setting-up errors is considered.

* *Department of Electrical and Electronic Engineering, Portsmouth Polytechnic, Portsmouth, PO1 3DJ.*
† *Department of Electrical Engineering, North East London Polytechnic, Dagenham, Essex.*

1. Introduction

In some branches of instrumentation, for example, electrometry, a single m.o.s.f.e.t. is used in the input stage of a d.c. amplifier. To minimize drift, it is desirable that the m.o.s.f.e.t. is operated at that drain current, I_{DS}, which gives a minimum variation of gate-source voltage, V_{GS}, with temperature, T.

The prediction of worst case circuit behaviour requires an analytical relationship between V_{GS} and T. This paper presents and discusses the implication of such a relationship with respect to device parameters and circuit performance limitations.

2. Analysis

In this contribution only the n-channel enhancement-mode devices are considered but the theory, with appropriate changes in the sign of certain parameters, is valid for the other m.o.s.f.e.t. variations.

For any m.o.s.f.e.t. one may write,

$$I_{DS} = f(V_{GS}, V_{DS}) \qquad \text{.....(1)}$$

where, in addition to the symbols already defined, V_{DS} is the drain-source voltage. If the device is operated in the 'pinch-off' region, then I_{DS} is not sensibly dependent on V_{DS} and equation (1) can be written as

$$I_{DS} = K\bar{\mu}(V_{GS} - V_T)^n \qquad \text{.....(2)}$$

where $\bar{\mu}$ is the effective mobility of majority carriers in the channel,

K is a constant related to oxide permittivity, ε_{ox}, channel width, W, channel length, L, and oxide thickness, t_{ox}, by the equation

$$K = \frac{\varepsilon_{ox} W}{2LT_{ox}}.$$

K is not significantly temperature dependent.

V_T is the projected gate-source threshold voltage

and n is the power law exponent, frequently taken as 2 but here uncommitted.

Equation (2) would appear to be the simplest non-trivial functional relationship[1] having general engineering application. If $g_{fs} = (dI_{DS}/dV_{GS})$, then n and V_T in equation (2) can be found from the best straight line fit to the experimental points on a plot of (I_{DS}/g_{fs}) against V_{GS}. The value of K could be found subsequently by using a particular I_{DS} in equation (2), but its value is not required in this analysis.

When the temperature varies and I_{DS} is maintained constant equation (2) yields,

$$\frac{dV_{GS}}{dT} = \frac{dV_T}{dT} - \frac{1}{n\bar{\mu}}\frac{d\bar{\mu}}{dT}(V_{GS} - V_T). \qquad \text{.....(3)}$$

Over a wide temperature range Wang *et al.*[2] have shown that (dV_T/dT) is a constant and we can write,

$$\frac{dV_T}{dT} = -a \qquad \text{.....(4)}$$

where a depends on channel doping concentration and

normally lies in the range 2–4 mV per degC. From equation (4),

$$V_T = V_{T0} - a\Delta T \qquad \text{.....(5)}$$

in which the threshold voltage is V_{T0} at a given reference temperature, T_0, and ΔT represents the deviation of the temperature from the reference value. Also

$$\frac{1}{\bar{\mu}}\frac{d\bar{\mu}}{dT} = -\frac{b}{T} \qquad \text{.....(6)}$$

where b, here left unspecified, is often taken as 1·5. Murphy *et al.*[3] quote four sources in support of equation (6).

From equations (3), (4), (5), (6),

$$\frac{dV_{GS}}{dT} = -a + \frac{b}{nT}[V_{GS} - (V_{T0} - a\Delta T)]. \qquad \text{.....(7)}$$

Equation (7) predicts a family of straight lines for (dV_{GS}/dT) when plotted against V_{GS}; the spacing along the V_{GS} axis depends on the threshold voltage and the slope of each line is (b/nT). Putting $b = 1{\cdot}5$, $n = 2$, $T = 300°K$, this gives 2·5 mV per degC per volt change in V_{GS}. In fact the results of Giralt *et al.*[4] give,

$$\left|\frac{d}{dV_{GS}}(dV_{GS}/dT)\right| = 2{\cdot}5 \text{ mV per degC per volt} \qquad \text{.....(8)}$$

in support of this. This same conclusion was reached by Cobbold.[5]

At low values of V_{GS}, (dV_{GS}/dT) is negative, while at higher values, (dV_{GS}/dT) is positive. The condition, $(dV_{GS}/dT) = 0$ is obtained from equation (7), giving,

$$V_{GS} = \left\{\frac{naT}{b}\right\} + V_T. \qquad \text{.....(9)}$$

Since (dV_{GS}/dT) is dependent on T, arranging to operate at $(dV_{GS}/dT) = 0$ does not ensure that $\Delta V_{GS} = 0$ for an increment ΔT in T. To find ΔV_{GS}, one may proceed as follows from equation (6),

$$\bar{\mu} = \bar{\mu}_0\left\{1 + \frac{\Delta T}{T_0}\right\}^{-b} \qquad \text{.....(10)}$$

where $\bar{\mu}_0$ is the mobility at the reference temperature. From equations (3), (5) and (10) we have at the reference temperature,

$$I_{DS} = K\bar{\mu}_0(V_{GS0} - V_{T0})^n \qquad \text{.....(11)}$$

where V_{GS0} is the reference value of V_{GS}. When an increment ΔT in temperature occurs, the gate source voltage increases by ΔV_{GS} and equation (11) becomes,

$$I_{DS} = K\bar{\mu}_0\left\{1 + \frac{\Delta T}{T_0}\right\}^{-b} \times \times[(V_{GS0} + \Delta V_{GS}) - (V_{T0} - a\Delta T)]^n. \qquad \text{.....(12)}$$

As the drain current is maintained constant, (11) and (12) may be equated. Solving for ΔV_{GS} gives,

$$\Delta V_{GS} = -a\Delta T - (V_{GS0} - V_{T0}) + (V_{GS0} - V_{T0}) \times \times\left\{1 + \frac{\Delta T}{T_0}\right\}^{(b/n)} - 1. \qquad \text{.....(13)}$$

Equation (13) is the required analytical relationship.

If only a small temperature excursion is to be considered, a binomial expansion of equation (13) can be employed.

$$\Delta V_{GS} = -a\Delta T - (V_{GS0} - V_{T0})\left[-\frac{b}{n}\frac{\Delta T}{T_0} - \frac{b}{n}\left(\frac{b}{n} - 1\right) \times \times \frac{\Delta T^2}{2T_0^2} - \frac{b}{n}\left(\frac{b}{n} - 1\right)\left(\frac{b}{n} - 2\right)\frac{T^3}{6T_0^3}, \ldots\right]. \qquad \text{.....(14)}$$

When the initial bias point satisfies the zero drift condition, $(V_{GS0} - V_{T0})$ may be eliminated, using equation (9), to give,

$$\Delta V_{GS} = \frac{a}{2}\left(\frac{b}{n} - 1\right)\frac{\Delta T^2}{T_0}\left[1 + \left(\frac{b}{n} - 2\right)\frac{\Delta T}{3T_0} + \ldots\right]. \qquad \text{.....(15)}$$

For $\Delta T \leqslant 10°C$ and typical values for b and n, the second and higher terms in the bracket of equation (15) are normally negligible. Thus to an accuracy of a few percent,

$$\Delta V_{GS} = \frac{a}{2}\left(\frac{b}{n} - 1\right)\frac{\Delta T^2}{T_0}. \qquad \text{.....(16)}$$

Equation (16) indicates a parabolic relationship between ΔV_{GS} and ΔT, with the vertex occurring at T_0. For $b = 1{\cdot}5$, $n = 2$, a maximum occurs at T_0 and $\Delta V_{GS} < 0$. If it were possible to have $b > n$, a minimum would occur at T_0 and $\Delta V_{GS} > 0$. Clearly, as $b/n \to 1$, $\Delta V_{GS} \to 0$. The possibility of varying b or n to meet such a requirement would therefore seem to merit investigation by device designers. In the case of p-channel enhancement mode devices, the polarity of the ΔV_{GS} values, discussed above for n-channel devices, are reversed, as $a < 0$.

Taking again, the typical values for b and n, namely, 1·5, and 2, respectively $a = 3$ mV per degC, $T_0 = 300°K$, $T = \pm 10$ degC, equation (16) indicates that $\Delta V_{GS} = 0{\cdot}125$ mV. This low figure would not generally be achievable in practice because of the difficulty and uncertainty in locating the zero drift point due to circuit tolerances, setting-up errors etc. If a small fractional error, ε, is made in setting I_{DS} at its zero drift point, the equivalent error, δ, in V_{GS}, is given by

$$\delta = \frac{\varepsilon I_{DS}}{g_{fs}}. \qquad \text{.....(17)}$$

Using equations (2) and (9),

$$g_{fs} = \frac{bI_{DS}}{aT} \qquad \text{.....(18)}$$

therefore,

$$\delta = \frac{\varepsilon aT}{b}. \qquad \text{.....(19)}$$

Thus V_{GS} is actually set at

$$V_{GS} = \left[V_{T0} + \frac{naT_0}{b}\right] + \delta \qquad \text{.....(20)}$$

or

$$V_{GS} = V_{T0} + \frac{naT_0}{b}\left(1 + \frac{\varepsilon}{n}\right). \qquad \text{.....(21)}$$

Substitution of this value of V_{GS} in equation (14) yields,

$$\Delta V_{GS} = -a\Delta T - \frac{naT_0}{b}\left(1+\frac{\varepsilon}{n}\right)\times \times\left[-\frac{b\,\Delta T}{n\,T_0}-\frac{b}{n}\left(\frac{b}{n}-1\right)\frac{\Delta T^2}{2T_0^2},\ldots\right] \quad \ldots\ldots(22)$$

or

$$\Delta V_{GS} \quad \frac{a}{2}\left(\frac{b}{n}-1\right)\left(1+\frac{\varepsilon}{n}\right)\frac{\Delta T^2}{T_0}+\frac{a\Delta T\varepsilon}{n}. \quad \ldots\ldots(23)$$

For $\varepsilon/n \ll 1$, equation (23) becomes,

$$\Delta V_{GS} = \frac{a}{2}\left(\frac{b}{n}-1\right)\frac{\Delta T^2}{T_0}+\frac{a\Delta T\varepsilon}{n} \quad \ldots\ldots(24)$$

where the second term on the right-hand side of equation (24) gives the additional drift resulting from non-ideal setting-up. With only a 1% setting error, i.e. $\varepsilon = 0{\cdot}01$, and the parameter values used above, the predicted $|\Delta V_{GS}|$ is increased by more than 100% from 0·125 mV to 0·275 mV for a ± 10 degC temperature variation.

3. Conclusions

The above analysis permits a prediction of the temperature drift in the gate-source voltage of a m.o.s.f.e.t. operated at constant drain current. Small errors in setting the drain current to the minimum drift condition can result in significantly increased drift figures.

4. References

1. Richer, I. and Middlebrook, R. D., 'Power-law-nature of field-effect transistor experimental characteristics', *Proc. Inst. Elect. Electronics Engrs*, **51**, pp. 1145–6, September 1963. (Letters).
2. Wang, R., Dunkley, J., De Massa, T. A. and Jelsma, L. F., 'Threshold voltage variations with temperature in mos transistors', *I.E.E.E. Trans. on Electron Devices*, **ED-18**, pp. 386–8, June 1971.
3. Murphy, N. St. J., Berz, F. and Flinn, I., 'Carrier mobility in silicon mosts', *Solid-State Electronics*, **12**, pp. 775–86, October 1969.
4. Giralt, G., Andre, B., Simonne, J. and Estere, D., 'Thermal drift of mos devices', *Electronics Letters*, **1**, pp. 185–6, September 1965.
5. Cobbold, R. S. C., 'Temperature effects in mos transistors', *Electronics Letters*, **2**, pp. 190–1, June 1966.

Manuscript first received by the Institution on 21st August 1972 and in final form on 2nd October 1972. (Short Contribution No. 160/CC152.)

STANDARD FREQUENCY TRANSMISSIONS—October 1972

(Communication from the National Physical Laboratory)

October 1972	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)		October 1972	Deviation from nominal frequency in parts in 10^{10} (24-hour mean centred on 0300 UT)			Relative phase readings in microseconds N.P.L.—Station (Readings at 1500 UT)	
	GBR 16 kHz	MSF 60 kHz	Droitwich 200 kHz	GBR 16 kHz	†MSF 60 kHz		GBR 16 kHz	MSF 60 kHz	Droitwich 200 kHz	GBR 16 kHz	†MSF 60 kHz
1	−0·1	−0·1	−0·1	627	614·7	17	0	0	0	634	611·6
2	−0·1	0	−0·1	628	614·6	18	0	0	0	634	612·0
3	+0·1	−0·1	0	627	607·9	19	0	0	0	634	611·8
4	−0·1	0	−0·1	628	608·3	20	0	0	0	634	612·0
5	−0·1	−0·1	0	629	608·9	21	−0·1	0	0	635	612·4
6	−0·1	−0·1	−0·1	630	609·4	22	0	−0·1	−0·1	635	613·8
7	−0·1	−0·1	−0·1	631	610·1	23	0	0	0	635	613·5
8	0	0	−0·1	631	609·8	24	0	+0·1	0	635	612·5
9	+0·1	0	−0·1	630	609·5	25	0	0	−0·1	635	612·5
10	−0·1	0	−0·1	631	609·0	26	−0·1	−0·1	−0·1	636	613·4
11	−0·1	−0·1	−0·1	632	609·6	27	0	−0·1	−0·1	636	614·1
12	−0·1	−0·1	−0·1	633	610·5	28	0	−0·1	−0·1	636	614·6
13	0	0	−0·1	633	610·8	29	0	0	−0·1	636	614·2
14	0	0	0	633	611·0	30	+0·1	0	−0·1	635	614·4
15	−0·1	0	0	634	611·1	31	−0·2	−0·1	−0·1	637	615·7
16	0	0	0	634	611·3						

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 AT Scale; $(AT_{NPL} - \text{Station}) = +468{\cdot}6$ at 1500 UT 31st December 1968.

Conference on Digital Processing of Signals in Communications

University of Technology, Loughborough 11th to 13th April 1972

The Conference was organized by the IERE with the association of the Electronics Division of the Institution of Electrical Engineers and the United Kingdom and Republic of Ireland Section of the Institute of Electrical and Electronics Engineers.

This article is based on reports prepared by Professor J. W. R. Griffiths, chairman of the Organizing Committee and other members of the Committee, namely Mr. M. S. Birkin, Dr. D. E. Pearson, Mr. L. K. Wheeler and Dr. V. J. Phillips.

The full list of papers read at the Conference was published in the March 1972 issue of *The Radio and Electronic Engineer* (page 540); the papers are available as IERE Conference Proceedings No. 23, which may be purchased from the Institution, price £7.50 post free.

This Conference was characterized by two features: of the 300 participants, nearly 75% stayed in the University of Loughborough's Halls of Residence and an unusually high proportion—23%—were from overseas. These factors contributed to the undoubted success of both formal and informal proceedings.

The technical scene was set in the opening, or keynote, address by Dr. R. W. Lucky, of Bell Telephone Laboratories. His immense authority on the subject enabled him to deal shrewdly and entertainingly with 'The promise and the problems' of Digital Signal Processing in Data Communications.

Computer Simulation

The opening session of the Conference comprised four papers on computer simulation techniques. The first paper, by R. Herman, N. G. Batty, M. Blench and D. L. Hedderly (Plessey Telecommunications Research) described a simulation program which had been developed to assist in the design of space satellite communication links. It was shown that with such a simulation it was possible to evaluate optimum parameters for a p.s.k. link and that the techniques involved could be of use in other fields.

Next R. J. Morrow and C. S. Warren (British Aircraft Corporation) described how a large hybrid analogue/digital computer had been used to analyse a complex communication system. The particular advantage claimed for the use of a hybrid computer was that a more realistic 'feel' of changes that have been made to system parameters can be given to the designer.

L. S. Moye and C. D. Nabavi (STL) and J. S. Bridle (Joint Speech Research Unit) then dealt in a joint paper with a program that could be used in the simulation of block diagrams of systems. It was shown how it was possible on relatively small computers to use such a program interactively for circuit design.

The fourth paper by C. C. Cock (STL) described a computer program for the study of how timing perturbations occurred in digital transmission systems. The paper showed by use of this program that this perturbation was caused by the mistuning of repeaters. This information could be used by the system designer to obtain a better understanding of the system constraints.

Digital Filters

Since the subject of digital filters could, and indeed has, provided sufficient material for a conference of its own, the papers in this session were mainly on aspects of digital filters which might relate to or have significance in the communication field. To some extent this meant that the papers themselves, apart from having the common factor of digital filters, did not bear too much relationship one to another.

The first paper by M. H. Ackroyd and F. Ghani (Loughborough University) discussed the use of filters in system identification, the system here could, of course, be a communication link. The method depended on the use of a short test sequence and discussed the optimum output processing device.

The evolution of a fully digital implementation of a correlation receiver was the subject of the next paper by A. C. Davies and M. M. Chawki (City University) and they indicated the simplifications that could be achieved by serial processing.

Delta modulation can be used instead of pulse code modulation to provide a binary representation of a signal and the paper by G. B. Lockhart (Leeds University) considered the problems and advantages of both non-recursive and recursive filters using delta modulation.

L. G. Cuthbert and P. R. Coward (Queen Mary College) then described the results of their work on the application of optimization techniques in the design of the tap coefficients in a finite duration impulse response digital filter.

Reduction of quantization errors in recursive digital filters was the subject of a paper by J. K. Stevenson (Hirst Research Centre). He showed that the input and output multipliers for optimum signal scaling may be set to integer powers of two and realized very simply by reordering the bits of binary coded signals or by using simple adders.

The paper by E. R. Broad and P. F. Adams (Post Office Research Department) showed how a template method of design (reminiscent of analogue template methods) could be used in the design of digital filters. Some practical examples were discussed.

J. D. Martin and J. Metcalfe (Bath University) discussed the problems of the realization of bandpass digital filters, and finally G. D. Cain (Polytechnic of Central London) showed how 'staircase' digital filters could be built up by using Hilbert transform elements.

Compression and Expansion

The Wednesday morning session of the conference was concerned with the compression and expansion of communication

signal bandwidth and coding techniques for accomplishing this. Of the eight papers in the session, six were concerned with picture coding and two with speech coding.

Professor T. S. Huang (Massachusetts Institute of Technology) discussed the requirements for the digital transmission of newspapers and magazines, and presented some calculations and experimental results relating to sampling density and the generation of moiré patterns. This was followed by a paper by C. E. Goodison (Meteorological Office, Bracknell) describing the current practice in weather chart transmission and proposals for an all-digital system. B. Wendland (AEG–Telefunken) talked about an adaptive coder for television signals with 5 to 1 compression over straight p.c.m., while B. G. Haskell (Bell Telephone Laboratories) described a technique for frame-to-frame encoding of videotelephone signals containing differential quantizing noise, which he illustrated with a short film.

In the post-coffee session, Professor P. A. Wintz (Purdue University) summarized the results and compression ratios obtainable with transform coding (about 3 or 4 to 1 compared with p.c.m.), with J. Poncin (CNET, Paris) adding some observations of his own on this type of coding. Two papers on speech processing concluded the session: J. S. Severwright (Loughborough University) presented results of studies of a speech-interruption technique with bit-rate reductions of 2 to 3, and J. R. James (Royal Military College of Science) described a limiter-encoder operating at 10 kbit/s.

Adaptive Systems

Five of the papers were concerned with serial digital transmission over audio bandwidth channels. R. J. Westcott (Post Office Research Station) provided a good resumé of the basic adaptive methods which can be employed and described the transmission characteristics to be encountered in the public switched telephone network. He then went on to describe an experimental adaptive modem, which largely made use of quantized feedback, designed to work at 4800 bit/s in such an environment. The two papers by R. C. Weston and F. W. Abbot (SRDE) and by R. L. Brewster (Aston University) both dealt with the aspect of substituting digital processing to the virtual exclusion of analogue methods, both delta and pulse code modulation being discussed. Weston also considered application to randomly dispersive channels, e.g. h.f. radio links. E. B. Stuttard (Racal-Milgo) completed the current scene by describing a commercially-produced adaptive modem developed for operation over leased point-to-point lines at up to 9000 bit/s. The equalizer functions wholly digitally and used m.o.s. stores in the delay line. J. D. Brownlie, (Post Office Research Station) provided an approach to the difficult problem of assessing the effects of decision errors during the adaptation process and the influence of signal statistics.

The other two papers continued into a broader field of adaptation. A. P. Clark and A. K. Mukherjee (Loughborough University) described a code-division multiplex system with adaptive detection and gave computed comparisons between serial and parallel systems. A. M. Rosie (Queen's University, Belfast) and collaborators from industry described an experimental adaptive digital filter capable of learning and distinguishing between two initially unknown signals occurring randomly in the presence of noise.

Due to the late withdrawal of a paper on the original programme, two short contributions were inserted: one by R. P. K. Galpin (Plessey Telecommunications Research) on adaptive equalization and the other by J. L. Shanks (Amoco Production, Oklahoma) on two-dimensional recursive filters.

Generally the session provided an excellent 'teach-in' for those on the fringe of the subject and a very useful forum for interchange of views between those actively concerned.

Signal Design

This final session occupied a whole day, the first papers in the morning being concerned with codes and with error detection and correction. Professor D. A. Bell (Hull University) described how the Bose-Chauduri-Hocquenghem algorithm can be applied to the design of non-binary error correcting codes in multi-level phase shift laying. A cellular array for digital scrambling, applicable to security systems such as are now being used in cash dispensers and for credit verification was put forward by K. J. Dean (Twickenham College of Technology).

An American paper, by N. P. Murarka (IIT Research Institute), showed how a signal processing technique based on the linear f.m. dispersion method of spectral analysis used in radar systems could be very effectively applied to achieve near optimum non-coherent detection of multi-channel f.d.m. f.s.k. signals.

Error correction in digital transmission systems employing scrambling clearly presents difficulties and G. G. Apple (Bell Telephone Laboratories) presented the theoretical bases for various conditions including the one-cell scramble case. Three authors from AEG-Telefunken, U. Haller, H. S. Matt and M. Progler, described a forward error-correction system developed for use in heavily disturbed data transmission, such as radio links, where a separate feedback channel is not readily available; two independent coding stages are used, one to deal with random errors due to a binomial distribution and the other to correct the remaining bursts. Error correction techniques include this type of forward error correction concept and also the automatic request for a repeat method and J. E. Blackwall (Racal-Milgo) described a modem incorporating both these correction systems.

The last paper of the morning session dealt with the design of a digital phase and frequency-sensitive detector having zero harmonic distortion at phase lock and infinitive captive range. It was contributed by L. F. Lind (Essex University) who also described its realization as three s.s.i. logic packages.

The first papers of the second part of this session were concerned with two aspects of delta-modulation. R. Steele and M. Passot (Loughborough University) presented an analysis of the delta-modulator in the slope-overload condition using Gaussian input signals; M. J. Hawksford (Essex University) and Professor J. E. Flood (Aston University) in their paper examined the various groups of patterns of pulses which occur in the transmitted waveforms and proposed some methods of adaptive delta-modulation.

The papers presented by authors from the Bell Laboratories, U.S.A. were both concerned with signal companding for pulse-code-modulation systems. A technique for the synthesis of circuits to convert digital signals from one compression law to another (particularly from μ-law to A-law) was described by P. W. Osborne and M. R. Aaron (BTL), with H. Kaneko (Nippon Electric, Japan). J. R. Sergo's (BTL) paper considered the generation of a companded signal by elimination of quantizing levels.

The final contribution was entitled 'A differential p.c.m. encoder for viewphone signals' and J. E. Thompson and G. A. Gerrard (British Post Office Research Department) showed a very interesting series of slides illustrating the quality of the pictures transmitted and the effects of varying various parameters of the system.

U.D.C. 518.5: 681.3

A Computer Algorithm for State Table Reduction

R. G. BENNETTS, B.Sc., M.Sc.,*
J. L. WASHINGTON, B.A., M.Sc.*
and
Professor D. W. LEWIN,
M.Sc., A.Inst.P., C.Eng., M.I.E.R.E.†

Based on a paper presented at the IEEE 'Eurocon 71' Conference held in Switzerland in October 1971.

SUMMARY:
As part of a large research programme on the use of computer aids for logic circuit design, it was required to provide an algorithm for the reduction of sequential circuit state tables (completely or incompletely specified). This paper discusses the theoretical problems of state reduction and demonstrates how a rapid solution may be obtained based on the determination and use of a closure function associated with each maximal or prime compatible set. The use of several heuristics ensures that a near-minimal solution to the subsequent closed-cover problem is always obtained, rather than the absolute minimum that is theoretically possible but computationally impracticable. This is in keeping with the overall design philosophy of producing a viable engineering design rather than the theoretical optimum usually dictated by switching theory. The algorithm has been programmed and the paper further discusses the data structures used and problems encountered in its implementation.

* *Department of Electronics, University of Southampton.*
† *Department of Electrical and Electronic Engineering, Brunel University, Uxbridge, Middlesex.*

1. Introduction

The use of switching theory as a methodology of design for sequential logic circuits is now firmly established. Briefly, the requirements of a sequential circuit can be represented by a *state diagram.* Such a diagram defines the internal state transitions and primary outputs when subjected to an input change and this represents a convenient starting point for designing the circuit. As a mathematical model, however, it is not so suitable and a tabular equivalent model can be constructed—this being referred to as the *initial state table.* In general, the number of state variables in this table may be reduced, and a systematic *state reduction* is implemented at this stage.

Each state variable in the *reduced state table* is now assigned a unique binary code (*state assignment*) and the resulting *fully-assigned state-table* is analysed with the objective of defining the *internal excitation equations* and *primary output equations.* These equations describe the physical circuit realization and may be implemented using standard logic elements.

Southampton University and Brunel University are currently collaborating in the development of a suite of programs known as CALD to assist in the design of combinational and sequential circuits, and these are based on the approaches suggested by switching theory.[1] This paper relates in particular to the reduction of the initial state table of a sequential circuit and proposes an algorithm that is considered suitable for implementation on a digital computer.

A knowledge of the fundamental papers of Paull and Unger[2] and Grasselli and Luccio[3] is assumed. In the interests of completeness however, some terms, that are used throughout the text, are defined in the next Section. It may be noted that some of the terms themselves represent a slight departure from the 'accepted' terminology of state reduction.

The departure is concerned with the usage of the words 'class' and 'set' and the present authors feel that the following terms are more strictly applicable to the concepts they define.

2. Definitions

Compatible Pair

Two states a and b are compatible if, starting from either, they cannot be distinguished by the subsequent output sequences when subjected to all finite input sequences. The compatibility relationship is written $a \sim b$.

Incompatible Pair

Two states are incompatible if they are not compatible.

Compatible Set—C-Set

A compatible set C_i is a set of states such that all possible pairs of states are themselves compatible,

e.g. if $a \sim b$, $a \sim c$ and $b \sim c$, then the set $C_i = \{a, b, c\}$ is a compatible set.

If the compatible set contains only one member, e.g. $\{a\}$, this is referred to as a *singleton.*

Incompatible Set—I-Set

An incompatible set is any set of states that is not a compatible set.

Maximal Compatible Set—MCS

A compatible set is a maximal compatible set if it is not a proper subset of any other compatible set.

Condition Class—CC

In general, the validity of a compatible set C_i is dependent upon the existence of a further set of compatible sets. Such a class P_i is referred to as the condition class for C_i and is written:

$$P_i = \{P_{i1}, P_{i2}, \ldots, P_{ij}, \ldots, P_{iq}\}$$

Note that if $P_i = \phi$, the null set, then this is a special case of compatibility, and the member states of C_i are said to be *equivalent* or *identical.*

Prime Compatible Set—PCS

A compatible set C_i is non-prime, if there exists another compatible set C_j such that

(i) $C_i \subseteq C_j$ *and*
(ii) $P_j \subseteq P_i$.

Otherwise C_i is a prime compatible set.

Note that a maximal compatible set is also prime.

Cover

A set of m compatible sets $C = \{C_1, C_2, \ldots, C_i, \ldots, C_m\}$ is said to cover a set of n states $S = \{a, b, \ldots, n\}$ if and only if $C_1 \cup C_2 \cup \ldots \cup C_m = S$.

Closure

Let $C = \{C_1, C_2, \ldots, C_i, \ldots, C_m\}$ be a set of compatible sets and let the condition class for C_i be given by:

$$P_i = \{P_{i1}, P_{i2}, \ldots, P_{ij}, \ldots, P_{iq}\}$$

If for all $1 \leq i \leq m$ and all $1 \leq j \leq q$, P_{ij} contains as a subset at least one of the member sets of C, then the set of compatible sets C is said to be closed.

3. The State Reduction Algorithm

Inherent in the processes involved in deriving state diagrams/tables is the possibility that sets of states will be compatible. If these can be identified and each set replaced by just one of the member states, a reduction in the total number of states can be effected. This reduction not only has implications with regard to the hardware realization but also can have a profound effect on the complex process of generating suitable fault diagnostic sequences.

The problem of state reduction resolves itself into two main areas: (i) the identification and derivation of the compatible sets (C-sets), and (ii) the selection of an optimal number that will provide a closed-covering on the initial set of states.

A major consideration in developing the algorithm is that the truly minimal version, although theoretically possible to derive, is not the main objective. Rather, an optimum reduced state table is all that is required since this may be achieved with less stringent requirements on both core store and time. This fits in with the overall CALD philosophy of producing a viable engineering design rather than the theoretical minimum usually dictated by switching theory.[1]

Throughout the paper the procedure will be illustrated using the state table originally described by Kella[4] and reproduced in Table 1.

Table 1. Initial state table

Present state	input/output x_1	x_2	x_3	x_4
a	*g*/1	—	—/0	—
b	—	*b*/—	*e*/—	—
c	—	—/1	*h*/—	*e*/—
d	—	—/0	—	*a*/1
e	*g*/0	*e*/—	*b*/1	—
f	—	*i*/—	—	*b*/0
g	*d*/—	*d*/—	—/0	*e*/—
h	—	*c*/—	—	*d*/1
i	*f*/1	*c*/0	—/0	—

The algorithm is based on the maximal compatible sets (MCSs) rather than the prime compatible sets (PCSs) for the following reasons:

(i) As the algorithm is not concerned with determining *the* minimum state table, the MCSs, representing a subset of the PCSs, reduces (in some cases drastically) the number of C-set candidates.

(ii) This reduced list of candidates allows for simple closure functions (C_f) associated with each MCS and consequently eases the problem of generating the closure tree (q.v.)†.

The major steps in the algorithm are as follows:

Step 1. Generation of the MCSs together with their associated closure functions.

Step 2. Identification of essential MCSs, if any.

Step 3. Starting with the essential MCS (or otherwise), generation of the closure tree to first termination.

Step 4. A check for complete cover with return to step 3, if necessary.

In programming this algorithm, the data structures cannot be naturally represented using arrays. It is more convenient to use a system where pointers may be set up between items of data to indicate an association. This can be done using a list processor, ring processor, or a more specialized processor for a complex data structure. Leaving the details of housekeeping to the processor frees the programmer from the time-consuming problems of data organization.

This approach, however, often incurs a heavy penalty, in that the organization absorbs large amounts of run time and, to a lesser extent, core store, in performing

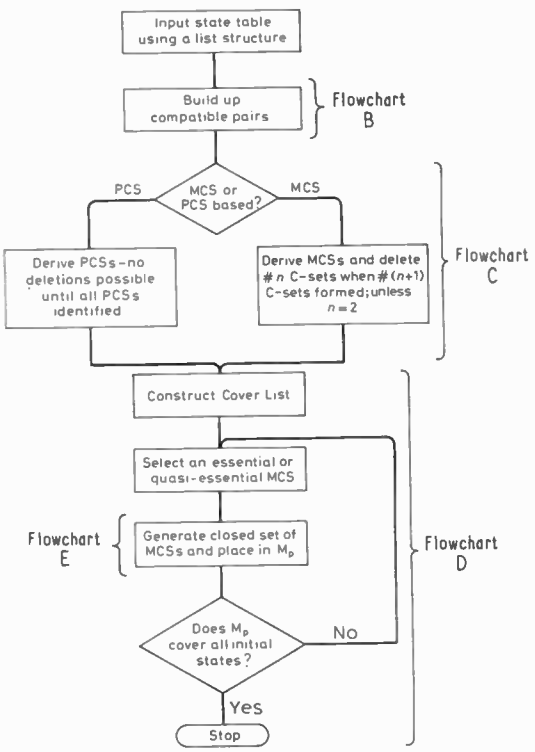
† In the program itself, provision has been made to calculate either the MCSs or PCSs. This allows two solutions to be obtained and enables a comparison between MCS- and PCS-based procedures. Apart from the considerations above, generation of PCSs can require considerably more core store than the MCS generation requirement.

activities which are more general than actually required for the particular application.

The initial problem studied for the CALD suite was that of combinational minimization, and considerable effort was devoted to producing a fast list processor embedded in *Fortran*. It was therefore natural to continue the commitment to this list processor, not only because it was available and well-proven, but also to ease the problems of interfacing between programs. This has a slight disadvantage in that some of the features, such as a fixed cell size, result in a sometimes inefficient use of store when this list processor is used on other problems.

The various steps through the algorithm will now be discussed and the individual flow charts are presented in Appendix 1. Flow chart A illustrates the overall procedure.

Flowchart A. Overall procedure for state reduction.



3.1 *Step 1: Generation of MCSs and the Closure Function C_f*

The generation of the MCSs is essentially the same as the technique, described by Bennetts,[5] for PCS derivation. In summary, it consists of:

(i) generating all compatible pairs together with their condition classes (CCs),

(ii) eliminating those having non-valid CCs,

(iii) constructing all n cardinality C-sets together with their CCs where $2 < n \leq n_{max}$, n_{max} being the natural limit,

(iv) terminating the list with all singleton classes,

(v) eliminating all non-maximal C-sets.

The complete list of all C-sets is shown in Table 2 and all those that are non-maximal are marked with a cross (×).

Table 2. Derivation of maximal C-sets and their condition classes

C-sets	Condition class	C-set	Condition class
×(*ab*)	(*φ*)	×(*bce*)	(*eh*)/3, (*bh*)/3 = (*beh*)/3
×(*ac*)	(*φ*)	×(*bcf*)	(*eh*)/3, (*bi*)/2, (*be*)/4
×(*ad*)	(*φ*)	×(*bcg*)	(*eh*)/3, (*bd*)/2
×(*af*)	(*φ*)	×(*bch*)	(*eh*)/3, (*de*)/4, [(*bc*)/2]
×(*ah*)	(*φ*)	×(*bde*)	(*φ*)
×(*ai*)	(*fg*)/1	×(*bdh*)	(*bc*)/2, (*ad*)/4
×(*bc*)	(*eh*)/3	×(*bdi*)	(*bc*)/2
×(*bd*)	(*φ*)	×(*beh*)	(*bc*)/2, (*ce*)/2 = (*bec*)/2
×(*be*)	(*φ*)	×(*bfg*)	(*bi*)/2, (*bd*)/2, (*di*)/2, (*be*)/4 = (*bdi*)/2, (*be*)/4
×(*bf*)	(*bi*)/2		
×(*bg*)	(*bd*)/2	×(*bhi*)	(*bc*)/2
×(*bh*)	(*bc*)/2	×(*ceh*)	(*bh*)/3, (*de*)/4, [(*ce*)/2]
×(*bi*)	(*bc*)/2	×(*cfg*)	(*be*)/4, (*di*)/2
×(*ce*)	(*bh*)/3	×(*deh*)	(*ad*)/4, (*ce*)/2
×(*cf*)	(*be*)/4	×(*dhi*)	(*ad*)/4
×(*cg*)	(*φ*)		
×(*ch*)	(*de*)/4	(*abcf*)	(*eh*)/3, (*bi*)/2, (*be*)/4
×(*de*)	(*φ*)	(*abch*)	(*eh*)/3, [(*be*)/2], (*de*)/4
×(*dh*)	(*ad*)/4	×(*abdh*)	(*bc*)/2, [(*ad*)/4]
×(*di*)	(*φ*)	×(*abdi*)	(*fg*)/1, (*bc*)/2
×(*eh*)	(*ce*)/2	×(*abhi*)	(*bc*)/2, (*fg*)/1
×(*fg*)	(*di*)/2, (*be*)/4	×(*adhi*)	(*fg*)/1, [(*ad*)/4]
×(*hi*)	(*φ*)	(*bceh*)	[(*beh*)/3], (*deh*)/4, [(*bec*)/2]
		(*bcfg*)	(*eh*)/3, (*be*)/4, (*bdi*)/2
×(*abc*)	(*eh*)/3	(*bdeh*)	[(*bc*)/2], (*ad*)/4, (*bce*)/2
×(*abd*)	(*φ*)	×(*bdhi*)	(*bc*)/2, (*ad*)/4
×(*abf*)	(*bi*)/2		
×(*abh*)	(*bc*)/2	(*abdhi*)	(*bc*)/2, (*fg*)/1, [(*ad*)/4]
×(*abi*)	(*fg*)/1, (*bc*)/2		
×(*acf*)	(*be*)/4	×(*a*)	(*φ*)
×(*ach*)	(*de*)/4	×(*b*)	(*φ*)
×(*adh*)	[(*ad*)/4] = (*φ*)	×(*c*)	(*φ*)
×(*adi*)	(*fg*)/1	×(*d*)	(*φ*)
×(*ahi*)	(*fg*)/1	×(*e*)	(*φ*)
		×(*f*)	(*φ*)
		×(*g*)	(*φ*)
		×(*h*)	(*φ*)
		×(*i*)	(*φ*)

3.1.1. Comments about the derivation of the maximal compatible sets

1. The starting point for constructing the list of all C-sets is the list of valid compatible pairs. This list is sub-partitioned into sets such that C-sets differ by one member only. In this way, two sets need only be tested for combination if they occur within the same sub-partition. Thus (*ab*) is only checked with (*ac*) to (*ai*). Note, however, that as the combination is to be non-transitive, if (*ab*) and (*ac*) exist, then (*bc*) must also exist in another sub-partition if (*abc*) is to be formed. Also, it is only necessary to check with those C-sets occurring *lower* than the 'master' C-set, i.e. (*ad*) would only be checked against (*af*), (*ah*) and (*ai*).

As C-sets of a higher cardinality are created, so this technique is extended and the sub-partitions again are

defined by those terms differing in one member only. This allows a short cut to be used when determining the existence of a larger C-set—namely that if $C_i = (a, b, c, d, \ldots, h, i)$ and also $C_j = (a, b, c, d, \ldots, h, j)$ $i \notin C_j$ and $j \notin C_i$ and if C_i and C_j are in the same sub-partition, then we can form a new set C_k where $C_k = C_i \triangle C_j = (i, j)$ in this case.† Note that the cardinality of C_k (denoted as $\#C_k$) must be 2 since C_i and C_j both occur in the same sub-partition. Now if C_k exists as a valid compatible pair, then it follows that C_i and C_j can be combined to form a larger C-set C_l $(= C_i \cup C_j)$. Connected with this, the corresponding condition class associated with C_l will consist of the union of those associated with C_i, C_j and C_k. These three C-sets will be sufficient to define the condition class regardless of the cardinality of C_l. (See Appendix 2 for a formal proof of these relationships.)

2. Associated with every C-set C_α is a condition class $P_\alpha = (P_{\alpha 1}, P_{\alpha 2}, \ldots)$. As the C-sets are derived, so the associated P_αs must also be constructed and modified. The following points should be borne in mind during this process:

(a) If P_α contains member sets $P_{\alpha i}$, $P_{\alpha j}$ both of which have the same input affiliation, then $P_{\alpha i}$ and $P_{\alpha j}$ must be replaced by $P_{\alpha i} \cup P_{\alpha j}$, e.g. for $C_\alpha = (bce)$, $P_\alpha = (eh)/3, (bh)/3$. Thus $P_\alpha \rightarrow (beh)/3$.

(b) If any member set of P_α, $P_{\alpha i} \subseteq C_\alpha$, then $P_{\alpha i}$ can be deleted from P_α, e.g. when $C_\alpha = (ceh)$ is formed, one of the P_α member sets is $(ce)/2$. Since $(ce) \subset (ceh)$, this is deleted from P_α. These deletions are shown thus: $[(ce)/2]$.

(c) The logical limit of such deletions is $P_\alpha \rightarrow \phi$, e.g. $C_\alpha = (adh)$, $P_\alpha = (ad)/4 \rightarrow \phi$.

The final list of maximal C-sets is shown in Table 3.

Table 3. Maximal compatible sets and closure functions

MCS	P_α^1	P_α^2	C_f (disjunctive) form)
C_1 (*abcf*)	(*eh*)(*bi*)(*be*)	(*eh*)(*abdhi*)(*be*)	$C_6C_3 + C_6C_6$
C_2 (*abch*)	(*eh*)(*de*)	[(*eh*)](*bdeh*)	C_5
C_3 (*bceh*)	(*deh*)	(*bdeh*)	C_5
C_4 (*bcfg*)	(*eh*)(*be*)(*bdi*)	(*eh*)(*be*)(*abdhi*)	$C_6C_3 + C_6C_5$
C_5 (*bdeh*)	(*ad*)(*bce*)	(*abdhi*)(*bceh*)	C_6C_3
C_6 (*abdhi*)	(*bc*)(*fg*)	[(*bc*)](*bcfg*)	C_4

In this list, the six MCSs are listed together with their P_αs, denoted here by P_α^1. Note that input affiliation is not now required. P_α^2 is an updated version of P_α^1, the modification occurring through the following rules:

(i) If for any P_α there exists $P_{\alpha i}$, where $P_{\alpha i}$ is an element of P_α and such that $P_{\alpha i} \subseteq C_\beta$, $C_\beta \neq C_\alpha$, and $P_{\alpha i}$ is not contained as a subset with any other MCS, then $P_{\alpha i}$ can be replaced by C_β.

† $C_i \triangle C_j = (C_i \backslash C_j) \cup (C_j \backslash C_i)$, where $\backslash$ is known as the difference operator and $\triangle$ is known as the symmetric difference operator.

The proof of this lies in realizing that if C_α is selected as a member of a cover set, then C_β will also be required to satisfy C_α's closure requirement.

e.g. for MCS (*bceh*), $P_\alpha^1 = (deh)$.

Now since $(deh) \subset (bdeh)$ alone, P_α^1 is updated to $P_\alpha^2 = (bdeh)$.

(ii) If, as a result of (i) above, $P_{\alpha i} \subseteq P_{\alpha j}$, or *vice versa*, then $P_{\alpha i}(P_{\alpha j})$ can be removed from P_α.

For example, for MCS (*abch*), $P_\alpha^1 = (eh)(de)$.

Initially $(eh) \subset (bceh)$ and $(bdeh)$, therefore (eh) is not modified. However, $(de) \subset (bdeh)$ alone and is therefore replaced by $(bdeh)$. Now (eh) becomes $\subset (bdeh)$ and can therefore be removed.

In programming this part of the algorithm, there are problems involving the storage of the intermediate C-sets. For MCS generation, this can be alleviated by noting that C-sets of size n can be deleted after they have been used to produce C-sets of size $(n+1)$. Nevertheless, there is still a considerable problem when large MCSs exist—if for instance, an MCS exists having a cardinality of n, then of the order of $n!/[(n/2)!]^2$ C-sets of cardinality $n/2$, will require simultaneous storage. In other words, the storage required roughly doubles for each increase in the size of the MCS.

For this reason, this part of the program is the most constraining. Even conserving storage by using a bit-oriented data structure does not significantly increase the program capability, since the required storage increases approximately at an exponential rate with n.

For this reason also, it is not possible to give an estimated upper bound on the size of the initial state table that can be handled in a given amount of core store. Obviously, this is dependent not so much on the initial number of states as on the internal structure of the state table itself.

The final stage in deriving the MCSs is to derive the associated closure function C_f. This is a convenient logical statement of the condition class P_α and is generated directly from the P_α^2 list. The procedure is summarized as follows:

(i) Assign C_i, $1 \leq i \leq n$ to each of the n MCSs.

(ii) For each C_i, determine the closure function by restating P_i^2 in terms of its conditional C_js.

For example, for $C_4 = (bcfg)$, $P_4^2 = (eh)(be)(abdhi)$.

Now $(eh) \subset (C_3 + C_5)$

$(be) \subset (C_3 + C_5)$

and $(abdhi) = C_6$.

Therefore the closure function for C_4 is given by

$$C_{f4} = (C_3 + C_5)(C_3 + C_5)C_6$$
$$= C_3C_6 + C_5C_6 \text{ in its disjunctive normal form.}$$

Note that each conjunct in C_{fi} represents an alternative set of MCSs necessary to satisfy the closure conditions imposed if C_i is selected. Thus if C_i is selected, at least *one* of the conjuncts must also be selected. This is not only a necessary condition, but is also sufficient. Inclusion of more than one conjunct is of no consequence.

The full set of closure functions associated with each MCS is shown in Table 3.

Implicit in the formation of a closure function is a conjunctive to disjunctive conversion (product-of-sum to sum-of-product). Such a process can require considerable core store and the approach adopted in the programmed version of the algorithm is to generate only the closure function when required, and then only to generate one of the terms in the disjunctive expression. The term that is generated may either be one of the smallest (although not necessarily *the* smallest) or may be modified to include existing chosen MCSs. This will be commented on further in the next section.

3.2. *Steps 2–4: Identification of Essential and Quasi-essential MCSs and Generation of a Closed Cover*

Theoretically, all possible combinations of MCSs that provide a cover on the initial set of states may be found by expanding a conjunctive cover expression into its disjunctive form—this being analogous to the algebraic solution of prime implicant tables. The problem is further complicated here however by the requirement for each cover set to be closed, and although the solution is theoretically soluble, its implementation is difficult—see, for instance, Grasselli and Luccio.[3]

The approach adopted here therefore is to generate a cover set that is closed by virtue of an integrated process whereby each additional cover set member is checked for closure and, if necessary, further members added. This results in the generation of a tree and the termination of any one branch indicates closure. Certain heuristics are applied that hopefully ensure that the number of branches is kept to a minimum, and also attempts an early termination. These are described later, but their application is assisted by reference to a cover list.

3.3. *The Cover List*

A cover list (CL) on the initial set of states can be constructed from the MCS list as shown in Table 4.

Table 4. The cover list (CL)

Initial state	Covered by
a	$C_6\ C_2\ C_1$
b	$C_6\ C_2\ C_3\ C_5\ C_1\ C_4$
c	$C_2\ C_3\ C_1\ C_4$
d	$C_6\ C_5$
e	$C_3\ C_5$
f	$C_1\ C_4$
g	C_4
h	$C_6\ C_2\ C_3\ C_5$
i	C_6

Some pertinent comments about forming the cover list are:

(i) The ordering of the cover terms for each state is important and is decided by two criteria.

Criterion 1: Cardinality of the MCS, i.e. if any state j is an element of C_m and C_n, and $\#C_m > \#C_n$, then C_m precedes C_n.

Criterion 2: Cardinality of the closure function conjuncts (if known). If $\#C_m = \#C_n$, then a priority may be assigned on the basis of the cardinality of the smallest conjunct of C_{fm} and C_{fn}, i.e. if the cardinality of the smallest conjunct of $C_{fm} < C_{fn}$, then C_m precedes C_n.

(This follows from the desire to keep the number of branches in the closure function tree to a minimum.)

As examples of this, see initial state *d* (Criterion 1) and initial state *e* (Criterion 2).

Note that the MCSs may be re-arranged initially using the two criteria above. Generation of the cover list is then a simple matter of, for each state, scanning down the list and selecting those MCSs that cover the state.

(ii) If an initial state is covered by one MCS alone, then this MCS is identified as *essential*.

(iii) If no essential MCS exists, then a quasi-essential MCS may be defined by identifying the initial state having the minimum number of covering MCSs and selecting the first one of these. Such a term is 'quasi-essential' for cover and by virtue of its being first will either provide maximum cover or will have the simplest conditional requirements, or both.

3.4. *Generation of a Closed Cover*

The procedure essentially consists of selecting one of the essential (or quasi-essential) MCS and attempting to satisfy its closure requirements. The result will be a closed set of MCS that may or may not provide full cover on the initial set of states.† If full cover is not achieved, the procedure is repeated, this time starting with an MCS that is essential or quasi-essential for at least one of the uncovered states.

The core of the procedure is the generation of a closed set of MCSs, starting from the essential MCS. This process is assisted by noting that the size of the terms in the closure function can be reduced if some of the MCSs have already been chosen. Furthermore, if these terms are not calculated until required, the calculation can be modified to take account of previously included MCSs.

This sort of strategy may result in a non-minimal solution but as absolute minimality is not a prime requisite, this is not considered to be disadvantageous—merely a limitation.

The procedure will be illustrated by deriving a closed cover from the closure function and cover list information in Tables 3 and 4.

Consider the cover list: C_4 and C_6 are both identified as essential MCSs and by definition, both must appear in the final closed-cover expression M. We will only explore one path at a time, however, and will select C_6 (since $\#C_6 > \#C_4$). Referring to the MCS/C_f table (Table 3), selection of $C_6 \rightarrow C_4$ anyway,‡ so the first

† These closed subsets are analogous to the irredundant prime closed sets of de Sarkar *et al.*[6]

‡ Here the arrow (→) is read as 'implies', i.e. $C_6 \rightarrow C_4$ means that C_6's validity rests on C_4's inclusion since C_4 is the condition class for C_6. Hence selection of C_6 implies also the selection of C_4.

Table 5. Results of applying the program to a number of examples

Example number	1		2		3		4		5		6		7	
Original states	9		14		15		9		22		29		8	
MCS or PCS-based?	M	P	M	P	M	P	M	P	M	P	M	P	M	P
Maximal or Prime C-sets generated	6	33	6	6	8	60	10	70	30	—	24	30	5	12
Number of C-sets selected	4	4	4	4	8	8	6	6	12	—	24	24	5	5
Theoretical lower bound	3		4		8		4		7		24		3	
Program run time (s)	35	56	39	58	47	90	54	124	251	—	142	142	42	49
Program run store (kbit)	2	4	3	3	2	6	3	12	14	—	4	4	3	3

Program core store = 12 kbits. All examples are incompletely specified.

level of generation of the tree is:

$$C_6 \rightarrow C_4$$

We must now check for closure on C_4 and also enter a partial closed-cover expression in M, namely $M_p = C_6C_4$. This allows a check mark to be placed against all initial states in CL covered by both C_6 and C_4, i.e. (*abcdfghi*). The closure check shows that $C_4 \rightarrow (C_6C_3 + C_6C_5)$ and we note that since C_6 is already in M_p, both conjuncts may be modified to $(C_3 + C_5)$. Alternatively, when calculating the C_4 closure function from the natural conjunctive form, all disjuncts containing C_6 may be removed. The next level of generation is:

$$C_6 \rightarrow C_4 \begin{cases} \rightarrow C_3 \\ \rightarrow C_5 \end{cases}$$

Of the two alternatives, the upper path is arbitrarily selected as the candidate and proceeding along the lines mentioned, the complete tree is:

$$C_6 \rightarrow C_4 \begin{cases} \rightarrow C_3 \rightarrow C_5 \rightarrow \phi \text{ (branch termination)} \\ \rightarrow C_5 \end{cases}$$

and the path from C_6 to ϕ is $M_p = C_6C_4C_3C_5$.

A scan down the cover list reveals that all states are covered and therefore $M = M_p$ in this case and the original nine-state table can be reduced to four states. Had full cover not been achieved a check would be made to identify essential terms not yet included and these would form the basis of another closure tree. If no essential terms were identified, then a pseudo-essential term would be defined based on the subset of initial states not yet covered. Working in this manner subsets of closed MCSs would be identified and added into M_p eventually to form the full cover M.

4. Concluding Remarks and Results

An algorithm for state table reduction based on the techniques of Paull and Unger[2] and Grasselli and Luccio[3] has been described. It is considerably modified, however, by algebraic considerations of individual C-set closure requirements which then allow certain heuristic procedures to be applied to the main closed-cover problem. The emphasis on the theoretical development of the algorithm, has been to produce a programmable technique that is aimed at deriving a 'good' engineering solution rather than the theoretical minimum.

The algorithm has been fully programmed for an ICL 1907 computer and may be used to either generate the maximal or prime C-sets. The subsequent problem of finding a closed cover is then common to both. In an attempt to examine the efficiency of the algorithm, the program has been used to reduce a number of initial state tables, each one being incompletely specified (the reduction of completely specified state tables may be regarded as trivial since all MCS/PCSs are required thereby removing the closed-cover problem). The results obtained by the program are shown in Table 5 and for all but one of the examples, two solutions obtained—one based on generating and selecting from the MCSs and the other based similarly on the PCSs.

With reference to these results, it is interesting to note that for all the examples for which dual solutions exist, the final number of states is the same for both the MCS- and PCS-based solution (the actual C-sets selected are different however). Also, the theoretical lower bound was obtained from the cardinality of the largest incompatible class set, although there is no guarantee that such a solution is possible.

Although only a small sample, this would seem to validate the initial statement that the added complexity of both generating and selecting from the prime C-sets is not, in general, justified. Indeed, for example number 5, the program is unable to generate all the prime C-sets within 32k of computer core store.† A solution for this particular example does however exist[7] and the number of prime C-sets is stated as being 261, enabling a reduction down to 9 states, compared with the 12-state solution based on maximal C-sets.

† This is an artificial constraint that has been placed on all programs that are part of the CALD suite.[1]

The computer run times, although comparable, are slightly longer for prime C-set based problems. This is mainly due to the extra time taken in identifying the prime C-sets, rather than in the subsequent selection of the closed cover. The main problem in fact, is in generating the prime C-sets within 32k of core store. As already noted, maximal C-set generation enables deletion of C-sets of size N after they have been used to produce C-sets of size $(N+1)$. This cannot be implemented in prime C-set generation as the complete set of C-sets of size N, where $1 \leq N \leq N_{max}$ is required before the prime C-sets may be identified.

On the basis of these comments, it would seem that the initial reasons for recommending maximal C-set rather than prime C-set based solutions are not so dominant as the fact that, irrespective of which base is chosen, the closed-cover process will tend to produce solutions containing a comparable number of states in a comparable time. The reason for recommending maximal C-set based solutions therefore, is primarily the fact that conceivably larger initial state tables can be accommodated and solved within a defined amount of computer core store.

One final comment regarding state table reduction in general will be useful. Most of the published literature in this area equates reduction in states with subsequent reduction in hardware. This hardware reduction takes two forms—the first being that the number of bistables required is equal to the number of state variables v, and this in turn is given by $\log_2 n$ (rounded up), where n = number of final states. As a consequence, bistable reduction is only achieved as n crosses a 2^v threshold, and these obviously become further apart as n increases. Even if the reduction does not cross such a threshold however, a second hardware reduction is usually achieved since any reduction in the number of states means that 'don't care' state assignments become available and these may be used to effect a more economical assignment—either through direct use as a 'don't care' in bistable input equation minimization or as supplementary assignments to eliminate critical races in asynchronous logic design.

Another important aspect of state reduction is in the state table analysis techniques for deriving testing sequences for sequential circuits.[8] These techniques are based on deriving an input sequence, the output response of which is capable of both verifying the existence of all states and also that all defined transitions between states can be made correctly. The complexity of generating such a sequence is directly related to the number of states and, as such, may potentially be the more important reason for the state reduction process.

5. References

1. Lewin, D. W., Purslow, E. and Bennetts, R. G., 'Computer Assisted Logic Design—the CALD System'. IEE Conference Publication No. 86, Computer Aided Design, April, 1972, pp. 343–50.
2. Paull, M. and Unger, S., 'Minimising the number of states in incompletely specified sequential switching functions', *Trans. Inst. Radio Engrs on Electronic Computers*, EC-8, pp. 356–7, July 1959.
3. Grasselli, A. and Luccio, F., 'A method for minimising the number of internal states in incompletely specified sequential networks', *Trans. Inst. Elect. Electronics Engrs on Electronic Computers*, EC-14, pp. 350–9, 1965.
4. Kella, J., 'State minimisation of incompletely specified sequential machines', *I.E.E.E. Trans. on Computers*, C-19, pp. 342–8, 1970.
5. Bennetts, R. G., 'An improved method of prime C-class derivation in the state reduction of sequential networks', *I.E.E.E. Trans. on Computers*, C-20, pp. 229–31, 1971.
6. de Sarkar, S., Basu, A. K. and Choudhury, A. K., 'Simplification of incompletely specified flow tables with the help of prime closed sets', *I.E.E.E. Trans. Computers*, C-18, pp. 953–6, 1969.
7. House, R. W., and Stevens, D. W., 'A new rule for reducing CC Tables', *I.E.E.E. Trans. on Computers*, C-19, pp. 1108–11, 1970.
8. Bennetts, R. G., and Lewin, D. W., 'Fault diagnosis of digital systems—a review', *Computer J.*, 14, pp. 199–206, May 1971.

6. Appendix 1
Flowcharts of the Algorithm

The following Flowcharts summarize the operation of the algorithm, both as an overall process (Flowchart A) and also describing the generation of compatible pairs, maximal or prime compatible sets and the optimal selection of a closed cover (Flowcharts B, C and D, respectively). The final Flowchart details the 'heuristic' processes involved in generating a near-minimal closed subset of MCSs or PCSs starting with an essential or quasi-essential MCS (PCS) and the partially specified solution so far obtained in M_p (which will equal ϕ initially).

7. Appendix 2
C-Set Construction and Non-transitivity

For an incompletely specified state table, the construction of C-sets of cardinality n, $n > 2$ is governed by the following theorems:

Theorem 1: A set of states is a compatible set if and only if every pair of states in that set is compatible. (This theorem is quoted from Paull and Unger.[2])

It follows from theorem 1 that for a C-set C_i of cardinality n to exist, there must exist $n(n-1)/2$ valid pairwise C-sets and these must be contained in C_i. Thus, construction of C-sets is strictly on a non-transitive basis.† It is the purpose of this Appendix to prove the following theorem and its associated lemma:

Theorem 2: If there exist two valid C-sets C_i and C_j having the same cardinality n, $n \geq 2$, such that

$C_i \triangle C_j = C_k$, $\triangle$ = symmetric difference operator

where $\#C_k = 2$,

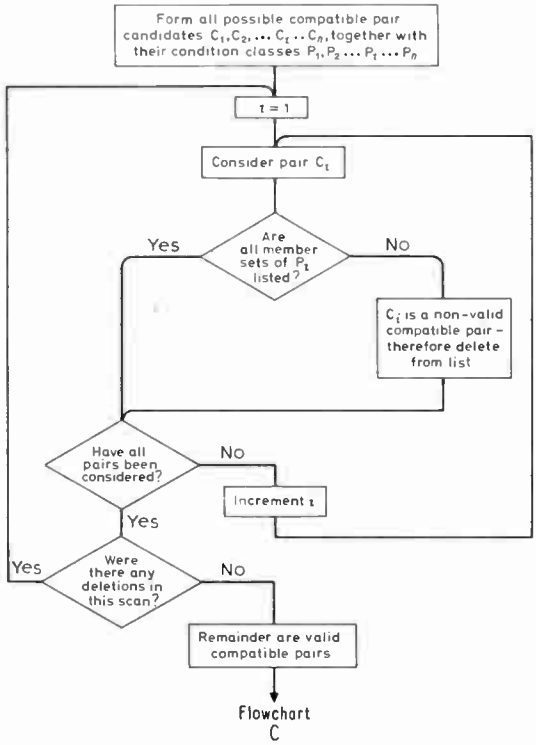
then provided C_k is a valid pairwise C-set, $C_l = C_i \cup C_j$ is also a valid C-set of cardinality $(n+1)$.

Lemma: In association with Theorem 2, if P_i, P_j and P_k are the condition classes of C_i, C_j and C_k respectively, then $P_l = P_i \cup P_j \cup P_k$ is the resulting condition class of C_l.

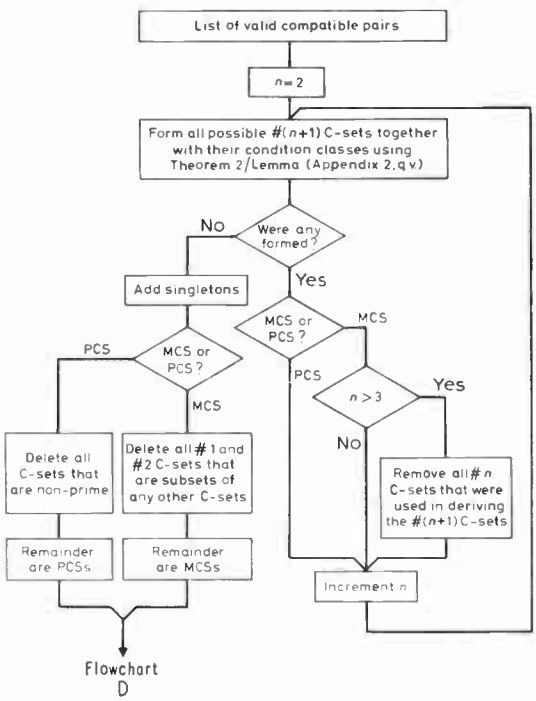
Proof: Both C_i and C_j will contain $n(n-1)/2$ valid pairwise C-sets (shortened to VPCS), but since $C_i \triangle C_j = C_k$, and $\#C_k = 2$, they will only specify $(n(n-1)/2+(n-1))$ *different* VPCSs. Consequently, the

† For completely specified state tables, a C-set may be constructed on a transitive basis and for C_i to exist ($\#C_i = n$) there need only exist $n/2$ (rounded up if n is odd) valid pairwise C-sets.

Flowchart B. Generation of compatible pairs.



Flowchart C. Generation of Maximal or Prime C-sets.



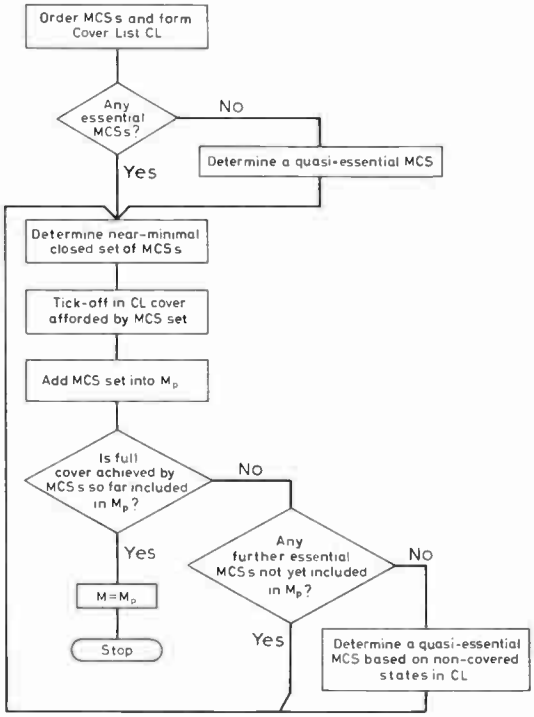
union of C_i and C_j will contain all different VPCSs specified by C_i and C_j plus one more (C_k), namely,

$$C_l = C_i \cup C_k \text{ will specify } \frac{n(n-1)}{2} + (n-1)+1$$

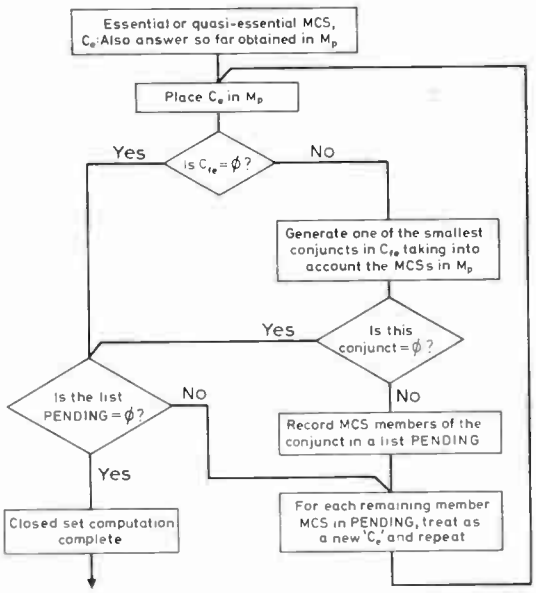
$$= \frac{(n+1)n}{2} \text{ different VPCSs.}$$

and since $\#C_l = (n+1)$, this number of VPCSs corresponds to the number defined by Theorem 1. Consequently C_l satisfies the validity condition.

Flowchart D. Generation of an optimal closed cover.



Flowchart E. Generation of a closed set of MCSs (or PCSs).



The condition class of C_l is strictly defined as being the union of the condition classes of all VPCSs contained in C_l. However, since C_i, C_j and C_k jointly specify all these VPCSs, P_i, P_j and P_k will jointly specify all condition classes and it follows that $P_l = P_i \cup P_j \cup P_k$ is the correct condition class for C_l.

Manuscript first received by the Institution on 24th May 1972 and in final form on 3rd July 1972. (Paper No. 1483/Comp 141.)