FOUNDED 1925
INCORPORATED BY
ROYAL CHARTER 1961

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

THE RADIO AND ELECTRONIC ENGINEER

The Journal of the Institution of Electronic and Radio Engineers

VOLUME 36 No. 4 OCTOBER 1968

'The World's Mine Oyster'

MAN's ability to project his talents on an international scale was frequently depicted in the works of William Shakespeare. Notwithstanding the technological revolution, the Bard's words* are, if anything, even more applicable to current affairs than they were in the sixteenth century.

Nowhere is this more evident than in the field of communications specifically and electronics generally. Modern communications technique has had a tremendous impact on the lives of all people. Industrial development in all countries is similarly creating widespread desire to understand and utilize the many applications of electronics in order to increase efficiency in both old and new industries, as well as to provide new tools in medical treatment and in such fields as oceanographic research.

These facts have been recognized by the establishment of the British National Electronics Council. Announcing that the Council was to be a focal point in the United Kingdom for studying the broad effect on national life of developments in the field of electronic science and engineering, the Minister of Technology stated: 'The rapid growth of the science and technology of electronics is one of the most striking phenomena of the present day. The continued application of electronics to more and more of the equipment we habitually use shows how our lives are being influenced to a growing extent by this new technique.'†

In particular the Minister, the Rt. Hon. Anthony Wedgwood Benn, emphasized the role of the National Electronics Council in securing '. . . the quick exploitation of new ideas and developments so that the benefits of advances in electronics can be enjoyed earlier than might otherwise be possible.' This can only be done by assisting the engineer to be aware of new discoveries and applications no matter where such developments occur. It was for this reason that N.E.C. had as its first major project the system now known as 'Selective Dissemination of Information'—a decisive contribution to overcoming the information explosion.

Thus the electronic and radio engineer is inevitably concerned with the development of electronics in every part of the world. This fact is recognized by the pages headed 'Conferences and Exhibitions' which have been a regular feature in *The Radio and Electronic Engineer* for the past decade and more. It is a 'current awareness' service which the *Journal* provides to its readers.

Some conferences may not, of course, contribute specifically to the fundamental knowledge of the electronic and radio engineer, but are invaluable to those engineers who have need to assess customer requirements—the subject of an earlier editorial in these columns.‡

For precisely the same reason 'Conferences and Exhibitions' frequently indicate joint activities between Societies and Institutions which enable engineers of other disciplines to understand more readily the contribution that electronics can make to their own efficiency, as well as enabling the electronic engineer to understand the problems of his fellow engineers in other specializations.

Whilst recognizing the political and economic problems attendant upon participation in many of these conferences, the venues sited in Great Britain, Germany, Italy, Rumania, U.S.A., Hungary, Holland, Canada, France, Sweden, and many other countries, demonstrate that indeed the electronic and radio engineer may truly say 'the world's mine oyster'.

G. D. C.

* From 'The Merry Wives of Windsor', II, ii, 2.
† *N.E.C. Review*, 3, No. 2, p. 29, October 1967.
‡ 'The Commercial Engineer', *The Radio and Electronic Engineer*, 35, No. 5, p. 249, May 1968.

INSTITUTION NOTICES

Colloquium on Fail-safe Techniques for High Reliability Computer Equipment

A half-day Colloquium on the above subject is being organized by the I.E.R.E. Computer Group Committee for inclusion in the joint programme of the I.E.R.E. and I.E.E. Computer Groups in February 1969.

Contributions are invited covering the techniques, both hardware and software, either in use or under development, to ensure fail-soft and/or fail-safe operation of high-reliability computer systems. Intending contributors are asked to submit short synopses—about 200 words—to the Secretary, Computer Group Committee, I.E.R.E., 8–9 Bedford Square, London, W.C.1, as soon as possible and in any case before 11th December 1968.

Full papers will not be required prior to the Colloquium; contributors are however invited to prepare papers for consideration for publication in the Institution's *Journal*.

Conference on Microwaves

The Institution of Electrical Engineers, together with the I.E.R.E. and the I.E.E.E., is organizing a European Microwave Conference which will be held in London from 8th to 12th September 1969. The Organizing Committee, which is under the chairmanship of Professor H. E. M. Barlow and on which the I.E.R.E. is represented by Dr. D. E. N. Davies, Professor W. A. Gambling, Mr. P. F. Mariner, Professor M. H. N. Potok, and Mr. T. W. Welch, would particularly welcome papers on the following:

Advances in Microwave Circuits: microwave networks and integrated circuits; filters and directional couplers; ferrite devices; delay lines and microwave acoustics; microwave and optical waveguide and waveguide components (excluding specific application to long distance telecommunications).

Microwave Antennas: inertia-less scanning adaptive aerials; active aerials.

Solid State Microwave Devices: low-noise receiving and amplifying devices; power sources; control and switching devices; bulk effect devices.

Synopses of papers (about 500 words) should be submitted by 5th January 1969 to the Joint Conference Secretariat, I.E.E.E., Savoy Place, London, W.C.2.

Publication of the Journal

The severe floods in South East England during September seriously affected the premises of The Whitefriars Press who print *The Radio and Electronic Engineer*. This means that publication of the *Journal* was delayed by about two weeks. Both the management and staff of the press have worked extremely hard to repair the damage, but there may still be some small delay in publication of this October issue.

The combined Programme Card for London Meetings and for East Anglian, Southern and Thames Valley Sections, was sent out with the September issue to members in the United Kingdom and although some of the meetings had taken place before the cards were in the hands of the printers, advance notice had fortunately been given of these meetings in the *Journal* or *Proceedings*, or in both publications.

Solid-State Physics

The Institute of Physics and The Physical Society has announced that its sixth annual conference on Solid-State Physics will take place at the University of Manchester from 7th to 9th January 1969. The object of the conference is to provide a forum covering the whole range of solid-state physics in order to promote the maximum interaction between specialists in different areas.

Introductory lectures on important fields in solid-state physics will be given by speakers from the United Kingdom and other European countries. Among the speakers will be:

S. Amelinox (Mol), J. R. A. Beale (Mullard), G. Chiarotti (Rome), J. Hubbard (A.E.R.E., Harwell), E. Kneller (Bochum), G. C. E. Low (A.E.R.E., Harwell), D. H. Martin (London), A. Seeger (Stuttgart) and J. Tauc (Prague).

Contributions on any topics of current interest in solid-state physics will be considered. The deadline for offers of papers has been set at 1st November 1968. Titles and brief outlines of 200 words should be sent before this date to the Conference Secretary, Professor S. F. Edwards, F.R.S., Department of Theoretical Physics, The Schuster Laboratory, The University, Manchester 13.

Further details and application forms are available now from the Meetings Officer, The Institute of Physics and the Physical Society, 47 Belgrave Square, London, S.W.1.

Change of Address

Members are requested not to delay in advising the Institution at 9 Bedford Square, London, W.C.1. immediately they change their address. A form for this purpose is frequently included in the back pages of the Journal. Because of delays in post, overseas members should advise their local secretary at the same time as they advise London.

Failure to notify the Secretary at 9 Bedford Square causes considerable delay in members receiving *Journals*, notices of meetings and other separate communication. In addition, extra postal charges fall on the Institution for *Journals* which have to be sent out a second time.

U.D.C. 621.314.5

Transistors in the Regulating Unit of Electronic Voltage Stabilizers

By

Professor

G. N. PATCHETT, B.Sc., Ph.D., C.Eng., F.I.E.E., F.I.E.R.E.†

Summary: Transistors may be used in both d.c. and a.c. voltage stabilizers and their use is considered in both applications although emphasis is placed on their use in a.c. voltage stabilizers. The operating conditions of transistors in a.c. stabilizers are different from their normal operating conditions in amplifiers and care is necessary to prevent damage to them. The operation of a transistor essentially as a variable resistor is considered. Transistors fed from auxiliary sources of power (either d.c. or a.c.) may absorb power or give out power and a number of circuits are considered including those using a complementary pair.

List of Symbols

V_z	Zener diode voltage
I_m	peak stabilizer load current
I	r.m.s. stabilizer load current
V_s	voltage of additional supply either d.c. or peak a.c.
V_{RU}	r.m.s. voltage of regulating unit
V_{cs}	control signal voltage either d.c. or r.m.s. a.c.
V_b	base bias voltage
W_{RU}	regulating unit power or change of regulating unit power
W_D	power dissipation in transistor (both transistors where two are used)
I_c	collector current
V_{ce}	collector-emitter voltage
I_b	base current

1. Introduction

A voltage stabilizer consists of a measuring unit, which detects any change of output voltage, and a regulating unit which makes the required correction in output voltage. The regulating unit may be placed in series with the load, or in parallel with the load together with a series impedance. Only the former case will be considered in this paper. Commonly the regulating unit uses valves or transistors but the paper is restricted to the use of the latter. Transistors may be used in both d.c. and a.c. voltage stabilizers and both cases are considered but emphasis is placed on the latter since little has been published previously. The paper is restricted to the use of dissipative elements and the use of transistors in non-dissipative circuits is not covered.

† School of Electrical and Electronic Engineering, University of Bradford.

2. Operation of Transistors in the Regulating Units of D.C. Voltage Stabilizers

The basic stabilizer circuit is shown in Fig. 1, the measuring unit (M.U.) feeding an appropriate signal to the regulating unit transistor TR1. This signal controls the transistor so that the voltage drop across it changes in such a direction as to maintain a constant or nearly constant voltage across the measuring unit and hence across the load. In all cases the p-n-p transistor may be replaced by an n-p-n transistor provided appropriate polarity changes are made.

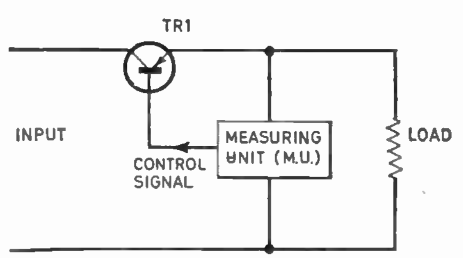


Fig. 1. Basic stabilizer circuit.

2.1. *Operating Conditions of the Transistor*

Under the worst operating conditions for the transistor, i.e. maximum input voltage and maximum output current (overload and short-circuit conditions are not considered) the following limitations of the transistor must not be exceeded:

(a) Maximum collector-emitter voltage.

(b) Maximum collector current.

(c) Maximum dissipation. This is the product of collector current and collector-emitter voltage. This dissipation must not be greater than that allowable under the highest ambient temperature conditions.

(d) Maximum voltage-current combination outside the secondary breakdown area.[1-3]

2.1.1. Transistors in parallel

Due to the above limitations the maximum output power from a stabilizer is limited if a single transistor is used. Transistors may be used in parallel to increase this power output. Due to the variation of base-emitter drop, load-sharing may not be good and small resistors may need to be connected in series with the emitters to improve the load-sharing. It is shown by McPherson[4] that for satisfactory load-sharing considerable power may be dissipated in these resistors and some compromise must be made between power lost in these resistors and accuracy of load sharing.

2.1.2. Transistors in series

An alternative method is the use of transistors in series. Each transistor now carries the full current and the voltage must be shared between them. Figure 2 shows a method of ensuring that the voltage is shared equally, ($R_1 = R_2 = R_3$). This series connection is generally only of value in relatively high-voltage stabilizers. An alternative series arrangement is shown in Fig. 3. Transistors TR1 and TR2 are fed with fixed base voltages such that appropriate voltages exist across the transistors. They act as emitter-followers or pre-regulators so that the voltage fed to the controlled transistor TR3 is approximately constant and hence improve the performance.

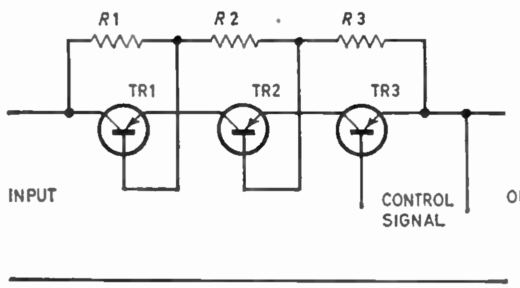


Fig. 2. Use of transistors in series.

2.1.3. Transistor in parallel with a resistor

The power output of a stabilizer may be increased by using a resistor in parallel with the transistor. When the transistor is cut off all the current flows and all the power is dissipated in the resistor. When the transistor is fully conducting practically all the current flows through the transistor and the power dissipated is small due to the small collector-emitter voltage. If the total current is constant, it can be shown that maximum dissipation occurs in the transistor when the currents in the transistor and resistor are equal. Under these conditions the transistor dissipation is only a quarter of the maximum regulating unit dissipation which occurs when the transistor is cut off. The most serious disadvantage of this arrangement is that the maximum available voltage drop across the regulating unit is proportional to the load current and hence, on no-load, the regulating unit is inoperative.

This difficulty may be overcome by the use of two transistors connected in one of the circuits shown in Fig. 4. These circuits use the principle of a transistor and resistor in parallel for large currents and a transistor and resistor in series for small currents. Consider the circuit of Fig. 4(a), the conditions being that of a large current and a control signal such that transistor TR2 is fully conducting. The battery voltage V_b is made greater than the drop across TR2 under these conditions and therefore TR1 is fully conducting, since its base is negative with respect to its emitter. Hence the total drop is small. If the control signal to TR2 is reduced (i.e. the base is made less negative with respect to the emitter) the voltage across TR2 rises, the base-emitter voltage of TR1 is reduced and the drop across TR1 increases, causing more current to flow in the resistor. Hence, the total voltage drop is increased. This action continues until TR1 is almost cut off and, with maximum load current, this voltage is arranged to be more than sufficient. Thus TR1 and R_1 operate as a resistor and transistor in parallel. For a constant current the maximum dissipation in TR1 is a quarter of the total maximum dissipation in the TR1–R_1 section of the regulating unit. The dissipation in TR2 is small since the voltage across it is small being less than the battery voltage V_b (say 2 V). When the current is small the voltage drop across R_1 may not be

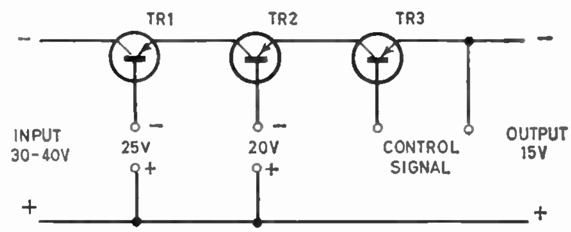


Fig. 3. Use of transistors TR1 and TR2 as pre-regulators.

sufficient and under these conditions the control signal is reduced, thus increasing the voltage across TR2. When this voltage exceeds the battery voltage V_b transistor TR1 is cut off. Thus R_1 and TR2 are in series and the regulating unit voltage drop can now be increased by increasing the voltage drop across TR2. The power dissipation in TR2 is small since the current is small. Maximum dissipation occurs in TR2 when the drop across TR2 and R_1 are equal, this then being a quarter of the maximum dissipation of this section of the regulating unit. The battery B is normally replaced by a Zener diode.

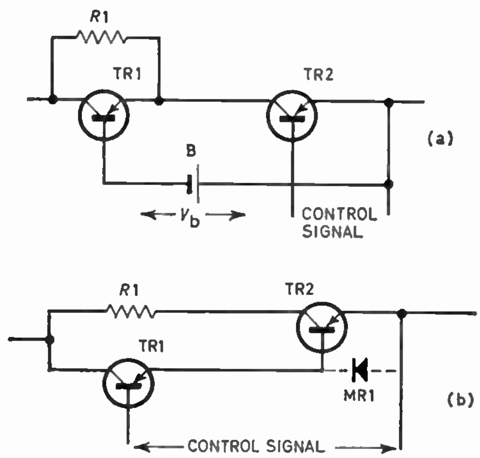


Fig. 4. Regulating units consisting of two transistors and a resistor for increased power.

In the circuit of Fig. 4(b) the action is similar. If TR1 is made heavily conducting then TR2 is also heavily conducting and essentially R_1 and TR1 are in parallel. Reducing the control signal to TR1 increases the drop across it forcing current to flow in R_1. During this period TR2 will be fully conducting as the base current of TR2 is the emitter current of TR1. As the control signal is reduced a point is reached where TR2 is no longer fully conducting and the voltage across TR2 rises and the circuit then consists essentially of TR2 and R_1 in series. This will only occur when the load current is so small that the drop across R_1 is not sufficient to maintain a constant output voltage from the stabilizer. In practice a rectifier (MR1) is added to bypass most of the emitter current of TR1 from TR2, in order that the base current of TR2 is not excessive. It is shown[4] that this is a special case of a more general arrangement of a number of transistors operating in sequence in this way, all having series resistors of different values. The advantages of dissipating large powers in resistors are: smaller transistors, smaller heat-sinks and greater reliability.

2.2. *Transistor Configuration*

The transistor in the regulating unit may be connected in one of three configurations: common-base, common-emitter or common-collector. Common-base is not a practical arrangement since the current gain is less than unity and the control signal current would be greater than the regulating unit current. When connected in common-emitter configuration (as in circuits described) there is a current gain β and only a small control signal current is required. Only a small control voltage is required which is equal to the base-emitter voltage, the input impedance is low and the transistor is essentially current-fed. If the control signal is applied between base and collector the transistor is in common-collector configuration. The current is as in the common-emitter configuration but the voltage is much greater, being the regulating unit voltage minus the base-emitter drop. Thus the input impedance is higher and the circuit tends towards being voltage-fed.

3. Operation of Transistors in the Regulating Units of A.C. Voltage Stabilizers

It will be assumed that the regulating unit is required for a 240 V 50 Hz stabilizer, the regulating unit being in series with the load. Valves have been used for this purpose[5–8] but little use has been made of transistors and hence the author considered a practical investigation should be made of the various methods of operation, these being more numerous than in the case of valves. If it is assumed that the maximum variation of input voltage is $\pm 4\%$, i.e. ± 10 V, then the maximum voltage across the regulating unit is 10 V, if equal buck and boost is assumed, or 20 V for a bucking voltage only. Since the voltage across the regulating unit is only a small fraction of the total voltage, the current in the regulating unit is settled by the load on the stabilizer and *not* by the voltage across the regulating unit. This is an unusual operating condition as the current is normally controlled by the input signal voltage. This is extremely important since, if the current is constant in this way, it is easy for the voltage across the regulating unit to rise to a value several times greater than normal as the voltage is determined by the impedance of the regulating unit, e.g. if the transistor is cut off then the supply voltage will appear across it. Since transistors will normally be selected to suit the normal voltage any such voltage-rise can be disastrous as regards the transistors. Thus some means must be used to limit the voltage-rise across the regulating unit. The most satisfactory way appears to be to connect a Zener diode in parallel with the transistor as shown in Fig. 5.

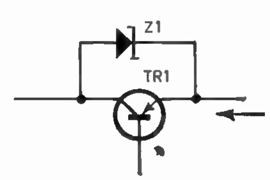


Fig. 5. Use of Zener diode to limit the voltage across the regulating unit.

When current flows in the direction shown, no current flows in the Zener diode until the voltage across the transistor reaches the Zener diode voltage. Any attempt to increase the voltage further causes current to flow in the diode, thus limiting the voltage-rise. If the transistor is cut off, all the current flows in the Zener diode, resulting in a square waveform of voltage across the diode and transistor. If current flows in the opposite direction it flows in the Zener diode and the voltage drop is small. The Zener diode, therefore, bypasses the current flow from the transistor. This is essential to prevent the occurrence of an excessive base-emitter voltage. If the power loss in the Zener diode during current-flow in the forward direction is neglected then the maximum power dissipated is $0{\cdot}45V_zI$ (see Appendix 1) where V_z is the Zener voltage and I is the r.m.s. regulating unit or stabilizer current. The Zener diode must be capable of dissipating this power and also of carrying the current I in both directions without damage. Suitable diodes are available for carrying currents up to several amperes.

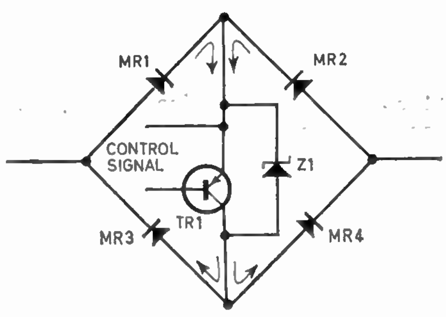
Valves must normally be fed through a step-up transformer as they are unsuited to operating on the low regulating unit voltage (10 to 20 V). However, transistors are very suited to this voltage and no transformer is necessary for this reason, but may be required for other reasons. Transistors may be operated in a number of ways but the methods of operation have been divided into two main classes: (i) those not requiring an additional supply, and (ii) those requiring a separate supply, either d.c. or a.c.

3.1. *Transistors Operated without any Additional Supply*

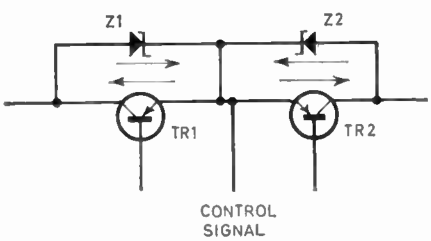
Since there is no additional power supply the regulating unit can only absorb power and therefore produce a bucking voltage. As a transistor is a unidirectional device, if symmetrical operation of the regulating unit is assumed, some means must be used to allow current to flow in both directions. Three basic arrangements are given in Fig. 6. Zener diodes are used in all cases to limit the voltage across the transistors. In Fig. 6(a) a single transistor is used and is fed by four rectifiers in a normal bridge circuit. Since current flows through the transistor during both half-cycles the dissipation will be twice that of each transistor in circuits shown in (b) and (c). In Fig. 6(b) is shown two transistors in series and with current flow in one direction it passes through Z1 (forward direction) and TR2, while in the other half-cycle it flows through Z2 and TR1. The parallel arrangement is shown at (c) where in one half-cycle current flows in TR1 and MR1 while in the other half-cycle it flows through TR2 and MR2. Due to the fact that a transistor is conducting when reverse voltage is applied and also the Zener diode is conducting when forward voltage is applied, rectifiers MR1 and MR2 must be added to prevent one parallel circuit shorting out the operation of the other. Circuit of Fig. 6(a) is used in Ref. 9. The present author prefers circuit (b) and this is used for the results that follow. The results apply equally well to circuits (a) and (c) provided allowance is made for the drop in the rectifiers.

3.1.1. Operation with d.c. control signal

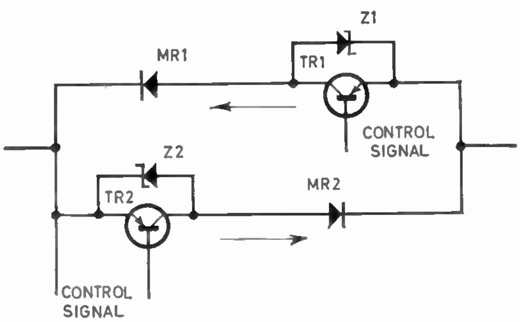
Common-emitter configuration: Typical characteristics of a transistor connected in common-emitter configuration are given in Fig. 7, where it will be seen that the slope resistance is high. Let the peak load current be I_m. It must be remembered that this current *must* flow through the regulating unit, either through the transistor or through the Zener diode. If the control signal is zero (i.e. $I_b = 0$) then, as the current rises at the start of the cycle, the voltage rises rapidly following the line AB until the Zener voltage V_z is reached. Further rise in current now causes current to flow in the Zener diode, following the line BC, the voltage remaining at a value V_z. The voltage across the regulating unit is almost of square waveform. If



(a) Bridge rectifier circuit



(b) Series connection



(c) Parallel connection

Fig. 6. Methods of connecting transistors in a.c. regulating units.

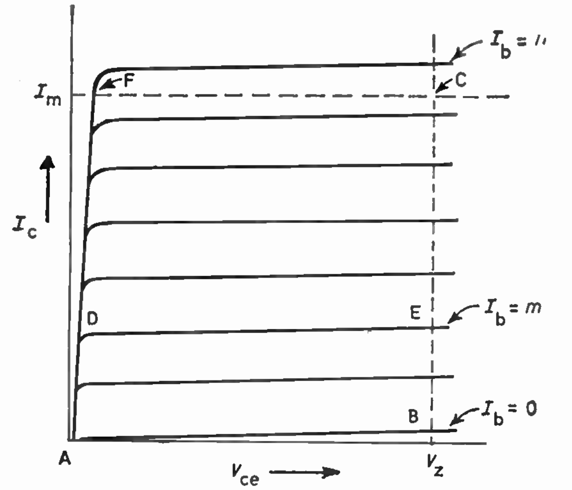


Fig. 7. Characteristics of a transistor in common-emitter configuration.

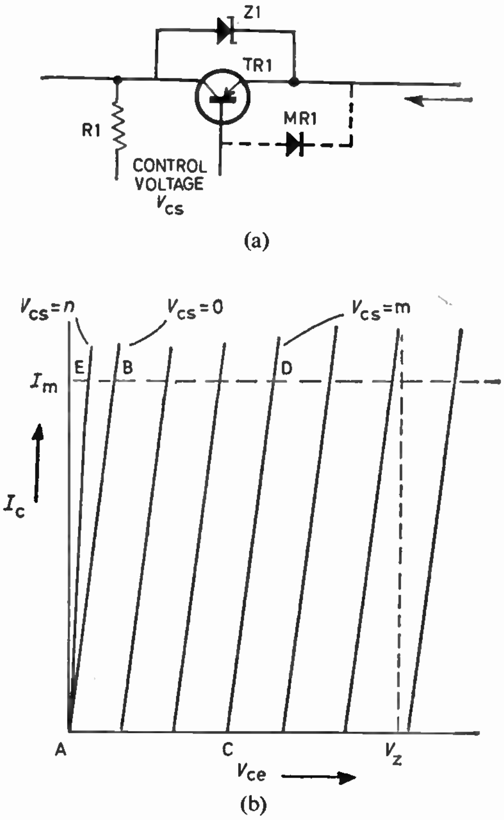


Fig. 8. (a) Common-collector circuit; (b) common-collector characteristics.

now a control signal current of $I_b = m$ flows, the operating line becomes ADEC. Thus there is little rise of voltage until the current reaches a value corresponding to D, when it rises rapidly to V_z and remains at this value. The waveform is therefore flat topped but narrower than in the previous case. With a sufficiently large control signal ($I_b = n$) the operating line is AF and there is little voltage across the regulating unit. Thus the effect of variation of control signal is to vary the width of the rectangular voltage waveform across the regulating unit. Due to the poor waveform the circuit is not satisfactory for many applications. The importance of distortion depends on the application.

Common-collector configuration: The circuit for one transistor is shown in Fig. 8(a), the control signal V_{cs} now being applied between collector and base via resistor R_1. Consider $V_{cs} = 0$. Any rise in collector-base voltage now causes an increased current to flow in R_1 and in the base, so tending to reduce the collector-base voltage. This negative feedback causes a change in the characteristics, the slope resistance being greatly reduced from, say, 50 kΩ to 50 Ω. A typical characteristic is shown in Fig. 8(b). The lower the value of R_1 the greater the negative feedback and the steeper the characteristics. When operated as a regulating unit with $V_{cs} = 0$ the operating line is AB, the voltage drop is small and approximately sinusoidal. If a control signal voltage $V_{cs} = m$ is applied making the base positive (i.e. opposing the negative feedback voltage) the characteristic is moved to the right and the operating line becomes ACD. The voltage drop is increased and the waveform is square in shape but with a curved top. Increasing V_{cs} increases the regulating unit voltage until it is limited by the Zener diode to a value V_z. If the control signal is reversed ($V_{cs} = n$) the operating line becomes AE and the voltage drop is reduced. A rectifier MR1 is added to prevent the application of excessive base-emitter voltage in the non-conducting direction. It should be noted that the control signal is now basically a voltage one (although a current will flow) and the peak regulating unit voltage is approximately equal to the control signal voltage. Again the waveform is poor but rather better than in the case of common emitter configuration.

The regulating unit voltage may also be varied by making $V_{cs} = 0$ and varying the value of R_1. In this case the slope of the characteristic is changed and the waveform of the voltage across the regulating unit is approximately sinusoidal. However, this method of operating the regulating unit is not usually convenient.

3.1.2. Operation with a.c. control signal

Common-emitter configuration: The alternating voltage control signal is now applied between the base and emitter, this voltage being in phase with the supply voltage and unity power factor load is assumed. With no control signal the operating line is ABC as in

Fig. 9 and as described earlier in Section 3.1.1. The waveform is square and of magnitude V_z. When a suitable alternating control signal current is applied, the operating line becomes one such as AD, due to the fact that the base current is proportional to regulating unit current. If equally-spaced characteristics are assumed then AD is a straight line and the voltage waveform is sinusoidal. In practice the characteristics tend to close up at large currents and hence line AD is curved and some distortion is present. The larger the control signal the steeper the line AD and the less the voltage across the regulating unit. Thus the voltage is reasonably sinusoidal until the voltage is limited by the Zener diode.

Common-collector configuration: The circuit is as in Fig. 8(a), the control signal voltage V_{cs} now being an alternating voltage.

As explained in Section 3.1.1, the transistor characteristics are now steep. When $V_{cs}=0$ the operating line becomes as shown in Fig. 10. If the control signal voltage is in a direction such as to add to the feedback voltage, the operating line is made steeper such as AB and AC. If the voltage is reversed then the slope is reduced and, for the condition of maximum voltage, the operating line is AD. A control signal greater than this results in a line such as AED and voltage limiting occurs due to the Zener diode. Provided the voltage is such that it is not limited by the Zener diode the waveform is good and this method of operation is superior to that described in Section 3.1.2. The peak dissipation under operating condition AD is high and this must not be sufficiently high to cause secondary breakdown, this breakdown occurring when a large current and voltage are used due to local heating of the junction.[1-3]

In the explanation of the operation it has been assumed that V_{cs} is in phase with the current. Since V_{cs} is normally in phase with the input voltage, this method of operation is only true for unity power factor. For other power factors the arrangement is not satisfactory as serious distortion occurs.

3.1.3. Operation in parallel with resistor

All the circuits described may be operated in parallel with a resistor but only common-collector circuits will be considered where the waveform is almost sinusoidal. As in the case of transistors operated on d.c. (see Section 2.1.3) the regulating unit power can be increased to approximately four times the allowable dissipation of the transistor but the arrangement is only suitable for approximately constant load conditions.

3.1.4. Limitations of transistors operated without any additional supply

Since there is no external source of energy the transistor can only absorb power and hence the voltage across the regulating unit can only be in one direction, namely, bucking. The allowable dissipation of the transistors must be equal to the power fed into the regulating unit unless a parallel resistor is used. If sinusoidal operation is considered, then this power will equal $\frac{1}{2}(V_{RU} \times I)$ for each transistor. Distortion often occurs if the power factor is not unity. At low power factors the power in the regulating unit becomes large, for a given difference between stabilizer input and output voltages. This is due to the fact that the voltage across the regulating unit is in phase with the current, and hence at a large angle to the input voltage and does not subtract directly as at unity power factor.

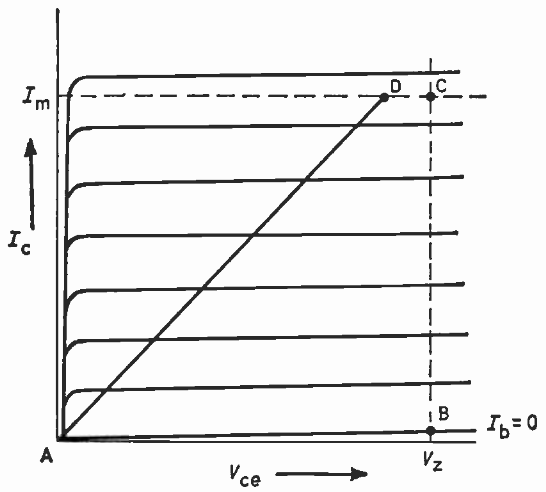


Fig. 9. Common-emitter characteristics and operating line with a.c. control signal.

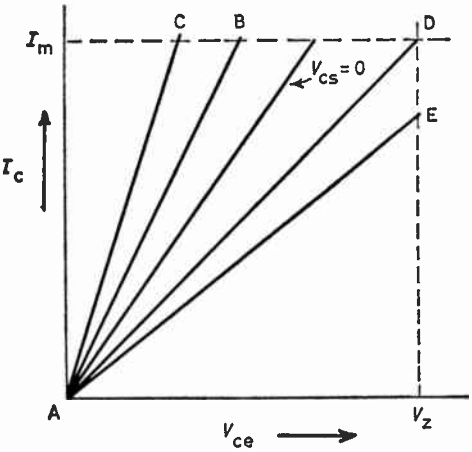


Fig. 10. Operation conditions with common-collector configuration and a.c. control signal.

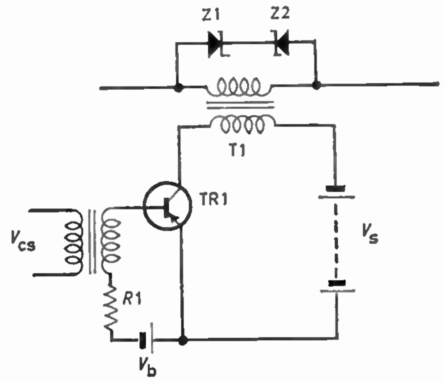


Fig. 11. Basic circuit for class-A operation using transformer; common-emitter configuration.

3.2. *Transistors Operated with an Additional D.C. Supply*

Since an external source of power is now available the regulating unit can give out power as well as absorb it and hence the voltage across it may be a boosting or bucking one.

3.2.1. Single-ended Class-A operation

Common-emitter configuration: The simplest arrangement is shown in Fig. 11 where a transformer is used in what appears a conventional circuit. Typical transistor characteristics are given in Fig. 12 where A is the static operating point. The variation of transistor current is settled by the load and hence the transistor must operate between, say, the two lines XY and PQ. If the base bias current $I_b = p$ and there is no control signal then the operating line is BC and the voltage drop is large, power is absorbed by the transistor and the regulating unit voltage is a bucking voltage. If now a control signal is applied in such a direction that an increase in the base current occurs with an increase of load current, the operating line moves anti-clockwise until it becomes a line such as DE and the voltage across the regulating unit is zero. This is explained elsewhere in connection with valves.[5, 6] Increase in the control signal will cause the operating line to move further clockwise to position FG and the transistor to give out power. Hence the voltage across the regulating unit is a boost voltage. A valve operates satisfactorily in this manner but the author found it impossible to get a transistor to give out power in this way. This is due to the characteristics which are almost horizontal. It will be seen from Fig. 12 that only a small control signal is required to cause a large change in the operating conditions. Thus, if the spacing of the characteristics is not the same this will upset the operating line and this is shown in an exaggerated form in Fig. 13. The characteristics above the static operating point A are drawn closer together than those below. Thus, when a control signal is applied so that the lower part of the operating line AF is in the correct direction for power output, the upper portion AG is not. Thus severe distortion occurs.

In practice the spacing changes gradually and hence the operating line is a curve. Due to this difficulty the circuit is useless as a regulating unit.

Common-collector configuration: With this configuration the characteristics are much steeper (as explained in Section 3.1.1) and more like those of a triode valve. In practice it was found that the transistor would operate satisfactorily but there are difficulties due to saturation of the transformer by the d.c. component of the collector current. Class-A operation is not efficient and hence this method of operation is not practical except for small powers, but it has been considered because the difficulties raised in connection with the common-emitter configuration in Section 3.2.1 appear again later.

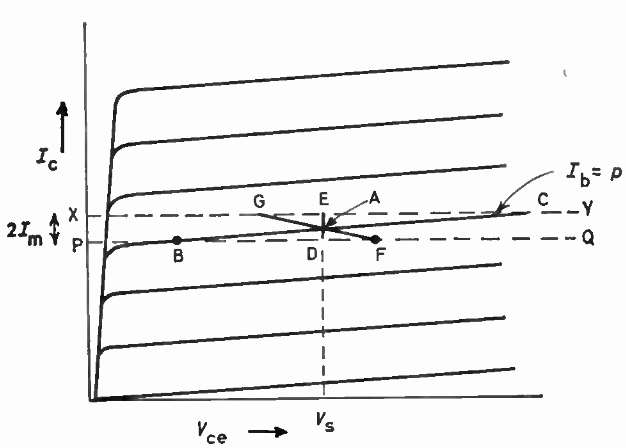


Fig. 12. Operation conditions of the circuit of Fig. 11.

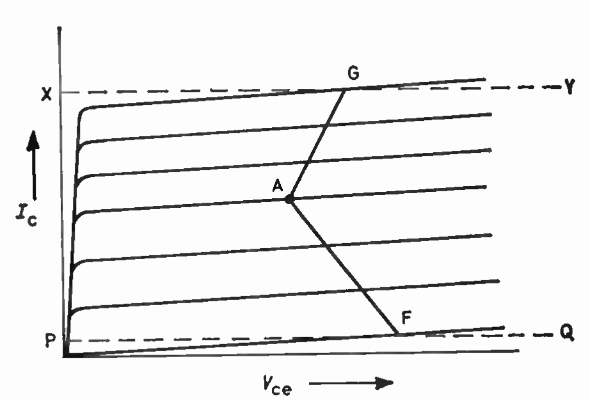


Fig. 13. Operating line for class-A common-emitter circuit showing serious distortion.

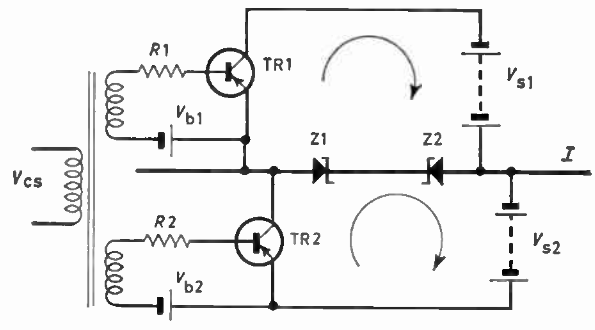


Fig. 14. Transformerless class-B common-emitter circuit.

3.2.2. Push-pull class-B operation—similar transistors

Common-emitter configuration: A transformer may be used as described in Section 3.2.1. However, a transformerless version is possible and the circuit is shown in Fig. 14. The author found no difficulty when the circuit was operated under class-A conditions, the circuit producing both a boosting and a bucking voltage. This is explained by the fact that the overall characteristic of the two transistors is linear. When the circuit was operated in class-B ($V_{b_1} = V_{b_2} = 0$) it was found that operation was satisfactory under bucking conditions but no power output could be obtained. The reason for this is similar to that for a single transistor operating in class-A. Although operated in push-pull, only one transistor is operating at a time and, due to the variations in the spacing of the transistor characteristics (which tend to be large in the case of power transistors), the operating line is very curved. A typical operating line is shown in Fig. 15, this being obtained by feeding the X-plates of a cathode-ray oscilloscope with the collector-emitter voltage and the Y-plates with a voltage proportional to collector current (the voltage across a low-value resistor). Hence the circuit is quite unsatisfactory.

Common-collector configuration: A circuit is given in Fig. 16 where it will be seen that the base is returned, not to the emitter, but to the opposite side of the Zener diodes, so that the regulating unit voltage is now applied between base and emitter. The magnitude of the resulting feedback current is determined by the values of the resistors R_1 and R_2 and the impedance of the control signal source. The voltage from the secondaries of the transformer T1 must now be rather greater than the voltage across the regulating unit. The characteristics of the transistor are now steep and some results are shown in Fig. 17, where it can be seen that the operating lines are very straight. Figure 18 shows that the waveforms are good. Slight cross-over distortion occurs and is particularly visible on the voltage taken with no control signal (Fig. 17(a)). It is important to note that this is taken with the c.r.o. at 100 times the sensitivity used for Fig. 17(b) and (c) and hence the distortion is small. This method of operation is very satisfactory and gives full buck and boost voltage with good waveform. The circuit will operate at all loads and all power factors. At power factors other than unity the operating line becomes an ellipse but the voltage across the regulating unit is in phase with the supply voltage and hence adds directly to, or subtracts directly from it. The power in the regulating unit relative to the transistor dissipation is calculated in Appendix 2 where it is shown that the maximum change of regulating unit power W_{RU} is:

$0{\cdot}88\,W_D$ where W_D is the maximum allowable dissipation of the transistors.

3.2.3. Class-B operation—complementary transistors

Common-emitter configuration: For the same reasons as described for the corresponding configuration in Section 3.2.2 this method of operation is unsatisfactory and will not be considered.

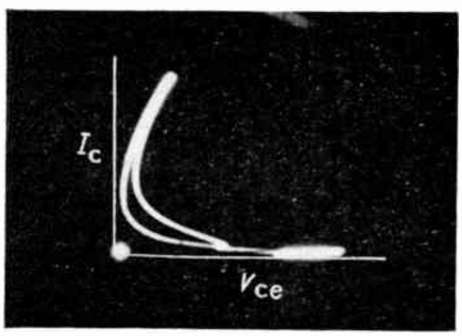


Fig. 15. Operating line for class-B common-emitter circuit under power output conditions.

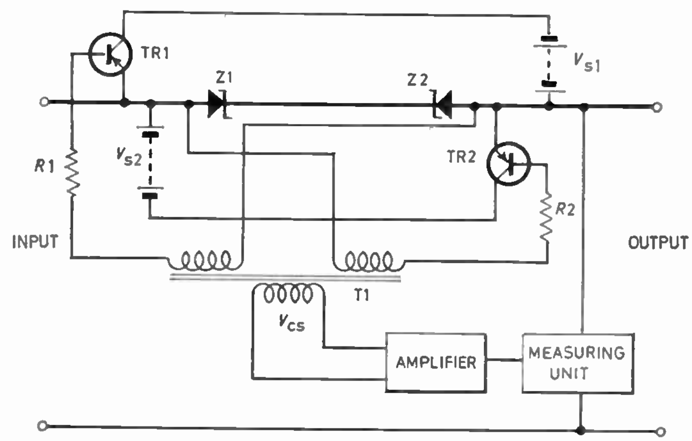
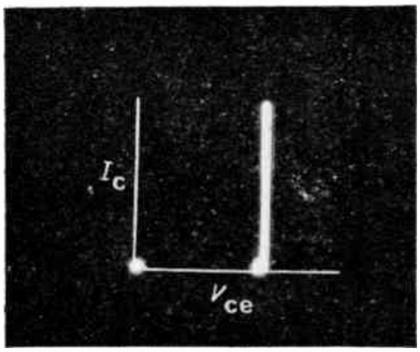
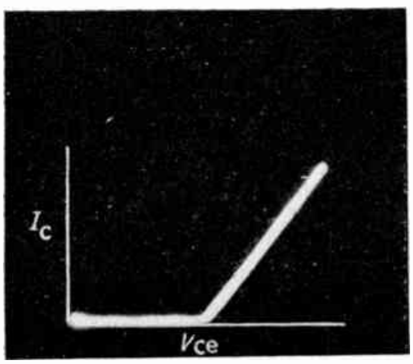


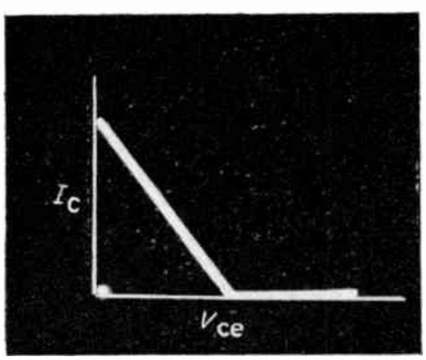
Fig. 16. Transformerless class-B common-collector circuit.



(a) No control voltage $V_{RU} = 0{\cdot}1$ V

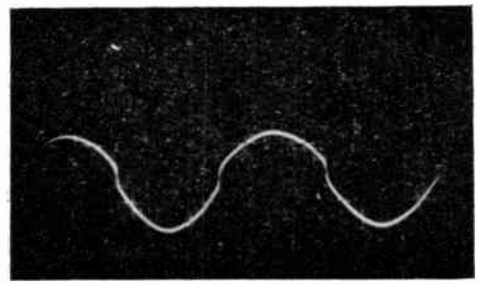


(b) Maximum bucking voltage condition $V_{RU} = 10$ V

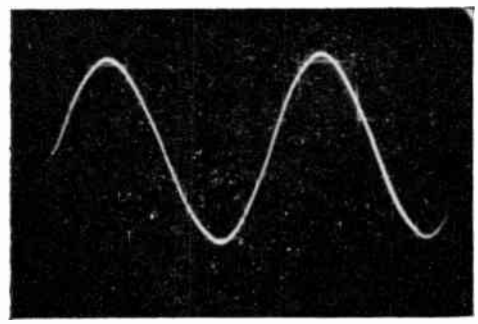


(c) Maximum boosting voltage condition $V_{RU} = 10$ V

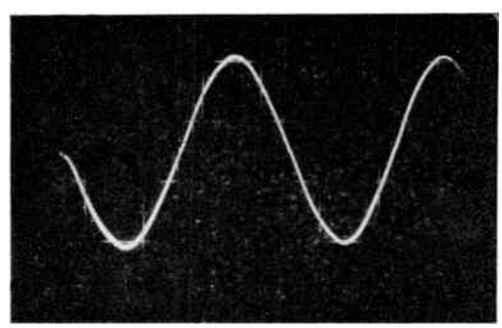
Fig. 17. Operating lines for class-B common-collector circuit. $V_{S1} = V_{S2} = 15$ V; $R_1 = R_2 = 0$; $V_z = 15$ V; $I = 1$ A.



(a) 0·1 V/cm;



(b) 10 V/cm;



(c) 10 V/cm.

Fig. 18. Waveforms corresponding to Fig. 17.

Common-collector configuration: The circuit is shown in Fig. 19 and has the advantage that only a single control voltage signal is required. As in the previous case the amount of negative feedback depends on the values of R_1 and R_2 (which are the same) and also on the impedance of the control signal source. The arrangement was found very satisfactory and results are shown in Fig. 20. Some slight distortion occurs due to the initial curvature of the characteristics but this is too small to be of importance. It may be removed by the application of a small standing bias to each transistor. The efficiency and transistor dissipation is as given in Appendix 2.

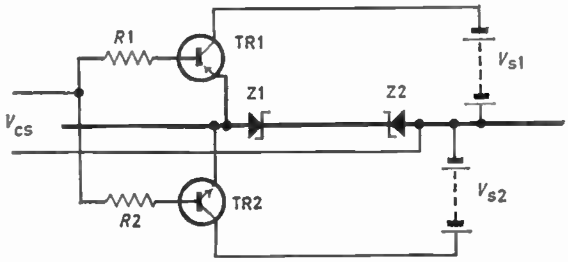
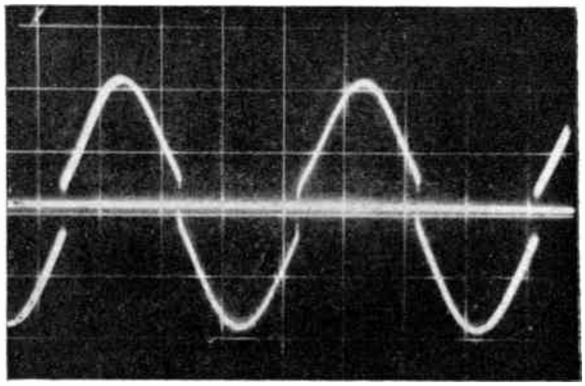
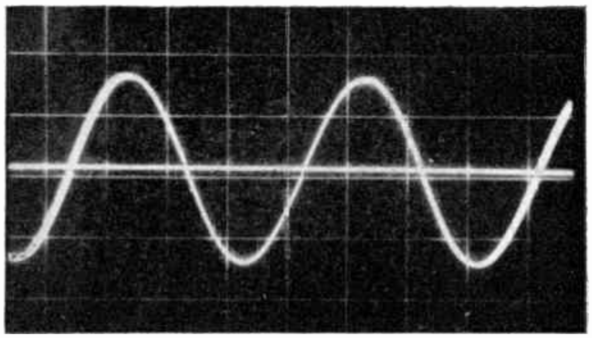


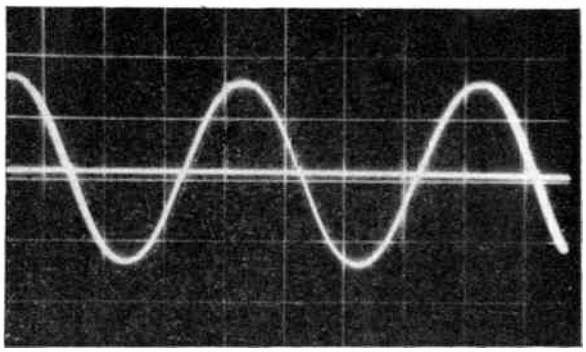
Fig. 19. Common-collector circuit using complementary transistors.



(a) No control signal $V_{RU} = 1$ V 1 V/cm



(b) Maximum bucking voltage $V_{RU} = 10$ V 10 V/cm



(c) Maximum boosting voltage $V_{RU} = 10$ V 10 V/cm

Fig. 20. Waveforms of the circuit of Fig. 19.
$V_{S1} = V_{S2} = 15$ V; $R_1 = R_2 = 120\ \Omega$; $V_z = 15$ V; $I = 1$ A.

3.3. *Transistors Operated with an Additional A.C. Supply*

Instead of using a steady d.c. supply as mentioned in Section 3.2, an a.c. supply may be used together with a rectifier, i.e. an unsmoothed half-wave supply. The arrangement is shown in Fig. 21 which uses transistors of similar type, but the principle can be used with a complementary pair. The advantage of an a.c. supply is that the ratio of change of regulating unit power to transistor dissipation is increased. When the regulating unit is giving out full power the transistor dissipation is small as the collector voltage is almost the supply voltage and when operating under bucking condition the dissipation is less than with a d.c. supply, as the mean collector-emitter voltage is less. It is shown in Appendix 3 that the maximum change of regulating unit power is almost equal to the maximum allowable dissipation of the transistors. Some results are shown in Fig. 22 where it can be seen that the distortion is small. This method of operation is cheaper than that using d.c. supplies as the cost of the additional supply is less. It has the disadvantage that serious distortion occurs if the power factor departs far from unity.

4. Conclusion

It is shown that transistors may be used in many ways in both d.c. and a.c. stabilizers. There are a number of methods of increasing the power of the regulating unit in d.c. stabilizers. When used in a.c. stabilizers there are many possible methods of operating transistors but not all are satisfactory. If operating essentially as a variable resistor, then the method of operation in Section 3.1.2 is the most useful as the waveform is good. If buck and boost voltages are required then common-collector operation is preferable either by using similar transistors (Section 3.2.2) or a complementary pair may be used (Section 3.2.3). Alternatively if some distortion is permissible then operation with an additional a.c. supply (Section 3.3) is suitable provided the power factor is near unity. Increased regulating unit power may be obtained by using transistors in parallel or in series in a similar way to that described in Sections 2.1.1 and 2.1.2. Means must normally be provided to prevent damage to the transistors by overload or short-circuit of the stabilizer terminals but discussion on this is outside the scope of this paper.

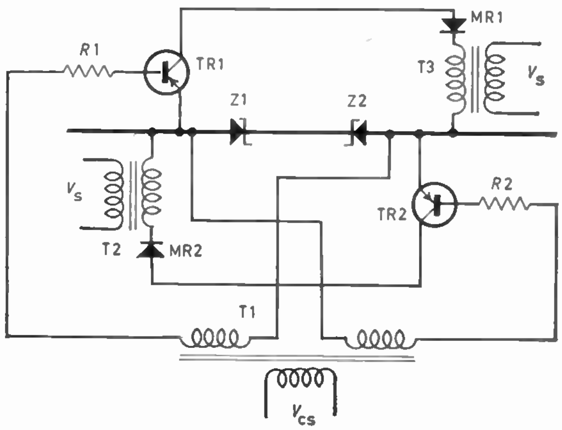
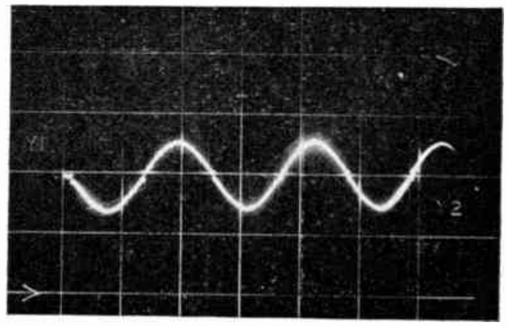


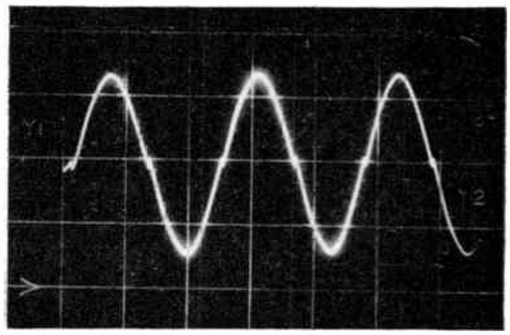
Fig. 21. Common-collector circuit using additional a.c. supply.

5. References

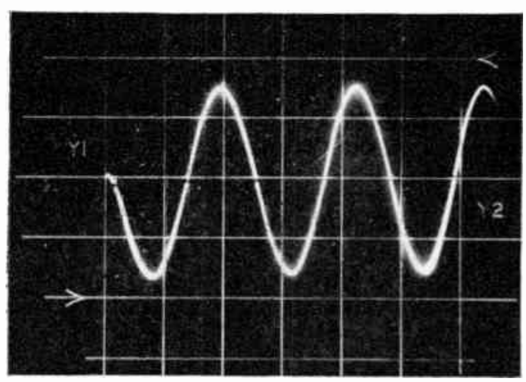
1. Hakim, E. B., 'An advancement in transistor second breakdown performance using molybdenum metalization', *Trans. Inst. Elect. Electronics Engrs on Reliability*, **R-17**, p. 45, March 1968.
2. Reich, B., 'New aspects of second breakdown in transistors', *Solid State Design*, **2**, April 1964.
3. Taue, J. and Abraham, A., 'Thermal breakdown in silicon p-n junctions', *Phys. Rev.* **108**, p. 936, November 1957.
4. McPherson, J. W., 'Regulator elements using transistors', *Electronic Engng*, **36**, p. 162, March 1964.
5. Patchett, G. N., 'Precision a.c. voltage stabilizers', *Electronic Engng*, **22**, pp. 371, 424, 470 and 499, September–December 1950.
6. Patchett, G. N., 'A new precision a.c. voltage stabilizer', *Proc. Instn Elect. Engrs*, **97**, Pt. II, p. 529, August 1950.
7. Clapp, C. W., 'Regulating a.c. with buck and boost amplifier', *Electronics*, **23**, p. 99, August 1950.
8. Dietest, H. W. and Hasler, M. F., 'Electronic voltage regulator for precise densitometric measurements', *J. Opt. Soc. Amer.*, **33**, p. 45, 1943.
9. Pim, J. A., 'Transistorized terminal equipment for submarine cables', *A.E.I. J. Telecommun.*, **1**, No. 2, p. 56, February 1967.



(a) No control signal 10 V/cm



(b) Maximum boosting voltage 10 V/cm



(c) Maximum bucking voltage 10 V/cm

Fig. 22. Waveforms of the circuit of Fig. 21.
$V_{S1} = V_{S2} = 10$ V; $R_1 = R_2 = 560\ \Omega$; $V_z = 15$ V; $I = 1$ A

6. Appendix 1
Power Dissipation in Zener Diode

Assume the transistor to be cut off so that all the current flows in the Zener diode and hence the voltage over the whole of the half-cycle is V_z. Dissipation on the other half-cycle is assumed negligible.

$$\text{Power} = \frac{1}{2\pi}\int_0^{\pi} iv\,\mathrm{d}\theta$$

where

$$v = V_z \quad \text{and} \quad i = I_m \sin\theta$$

$$\text{Power} = \frac{1}{2\pi}\int_0^{\pi} V_z I_m \sin\theta\,\mathrm{d}\theta$$

$$= 0{\cdot}32 V_z I_m$$

$$= 0{\cdot}45 V_z I$$

where I is the r.m.s. regulating unit current

$$= \frac{1}{\sqrt{2}} I_m$$

7. Appendix 2

Transistor Dissipation with D.C. Supply

In Fig. 23 the static operating point is A. The operating lines are as follows:

no control signal	AB
maximum bucking voltage	AC
maximum ideal boosting voltage	AD
maximum practical boosting voltage	AE

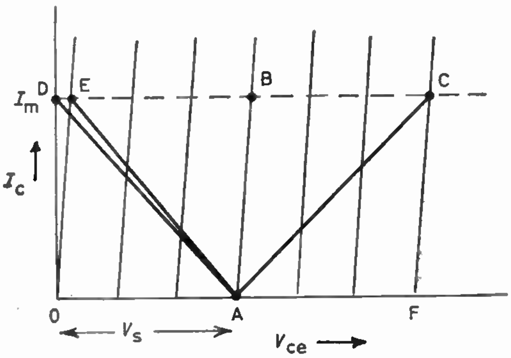


Fig. 23. Operating conditions of transistor with equal buck and boost voltages.

If equal buck and boost is assumed then OF $= 2 \times$ OA and the peak regulating unit voltage is the supply voltage $V_s (= \text{OA})$.

Assume unity power factor.

The instantaneous current

$$i = I_m \sin\theta$$

and the instantaneous voltage

$$v = V_s \pm V_s \sin\theta$$

Therefore the maximum dissipation of both transistors W_D

$$= \frac{1}{\pi}\int_0^{\pi} (I_m \sin\theta)(V_s + V_s \sin\theta)\,d\theta$$

$$= 1{\cdot}14 V_s I_m$$

But

$$V_s = V_{RU}\sqrt{2} \quad \text{and} \quad I_m = I\sqrt{2}$$

and the maximum dissipation

$$W_D = 2 \times 1{\cdot}14 V_{RU} I$$

$$= 2{\cdot}28 V_{RU} I \quad \text{and} \quad V_{RU} I = \frac{W_D}{2{\cdot}28} \quad \ldots\ldots(1)$$

Since the regulating unit gives equal buck and boost voltage, the maximum variation of regulating unit power W_{RU} is $2V_{RU}I$.

Substitution from eqn. (1) above for $V_{RU}I$ gives

$$W_{RU} = 2\frac{W_D}{2{\cdot}28} = 0{\cdot}88 W_D$$

In practice, the dissipation will be slightly less than this as the theoretical operating line AD has been used rather than the practical line AE.

8. Appendix 3

Transistor Dissipation with A.C. Supply

Assume peak auxiliary supply voltage V_s to be equal to maximum peak regulating unit voltage.

Under bucking conditions and unity power factor, the instantaneous current

$$i = I_m \sin\theta$$

and the instantaneous voltage

$$v = \text{supply voltage} + \text{regulating unit voltage}$$

$$= V_s \sin\theta + V_s \sin\theta$$

$$= 2V_s \sin\theta$$

Thus maximum dissipation of both transistors, W_D,

$$= \frac{1}{\pi}\int_0^{\pi} (I_m \sin\theta)(2V_s \sin\theta)\,d\theta$$

$$= V_s I_m$$

But

$$V_s = V_{RU}\sqrt{2} \quad \text{and} \quad I_m = I\sqrt{2}$$

and the maximum dissipation

$$W_D = 2V_{RU} I \quad \text{and} \quad V_{RU} I = \frac{W_D}{2} \quad \ldots\ldots(2)$$

Since the regulating unit gives equal buck and boost voltages the maximum change of regulating unit power $W_{RU} = 2V_{RU}I$.

Substitution from eqn. (2) gives

$$W_{RU} = 2\frac{W_D}{2} = W_D$$

In practice the dissipation will be slightly less than this as there will be a small voltage drop in the transistor which has not been allowed for.

Manuscript first received by the Institution on 21st February 1968 and in final form on 3rd July 1968. (Paper No. 1212/CC21.)

Letter to the Editor

Superdirective Arrays

SIR,

We agree with Dr. Longstaff† that it is not necessary to have active elements to produce a superdirective array. We have been working for several years on a type of multi-beam antenna (variously called the Maxson or Blass array) and have suggested that this beam-forming network could be used to produce a superdirective array.

The Maxson multi-beam antenna is a device for simultaneously producing several beams from a single aperture in the manner shown in Fig. A. The input/output terminal ports are all connected to the array of radiators through a matrix of directional cross-couplers, which couple power only in the directions indicated by the arrows. The couplers in feeder arm 1 control the aperture amplitude distribution for beam 1, those in feeder arm 2 for beam 2 etc., whilst the set of phase inserts 1 control the constant phase increment between radiators for beam 1, and the set of phase inserts 2 act similarly for beam 2 and so on. In this way each beam can be controlled both in beamwidth or side-lobe level and in its angular direction relative to the aperture. As with any other type of multi-beam antenna the beams have to be orthogonally spaced in phase (ϕ) space for the network to maintain the independence between the beams with no net interaction energy. In other words, when the beams are orthogonally spaced, one beam is identified with one terminal port alone there being no energy transfer between that beam, other terminal ports, other beams, or the terminating loads in the network.[1] The angular position θ of a main beam, relative to the normal to the aperture, is determined by the relation

$$\dot{\phi} = \frac{2\pi d}{\lambda} \sin \theta$$

where ϕ is the constant phase increment between the radiators, d is their distance apart, and λ is the wavelength.

Now it was shown by Woodward[2] that any far-field pattern formed in real space from an aperture can be resolved into an angular spectrum of $(\sin \phi)/\phi$ field patterns. If we synthesize a field pattern from the orthogonally-spaced $(\sin \phi)/\phi$ patterns, which we believe may be formed quite independently by the Maxson matrix, we can define the field pattern at n points by using an n terminal port matrix to feed n radiators. If the n radiating elements fed by the matrix are at $\lambda/2$ spacing the component $(\sin \phi)/\phi$ patterns available are just those of the real space. However, Woodward showed that to obtain a superdirective pattern it is necessary to use also the $(\sin \phi)/\phi$ patterns whose main beams lie in imaginary space ($\sin \theta \geqslant 1$) and that it is the real-space side-lobes of these imaginary beams which impose superdirectivity on the pattern. We therefore must use our n-port matrix to feed n radiating elements at less than $\lambda/2$ spacing, in order to have available sufficient of Woodward's imaginary beams. We see no reason why the n terminal ports of the matrix

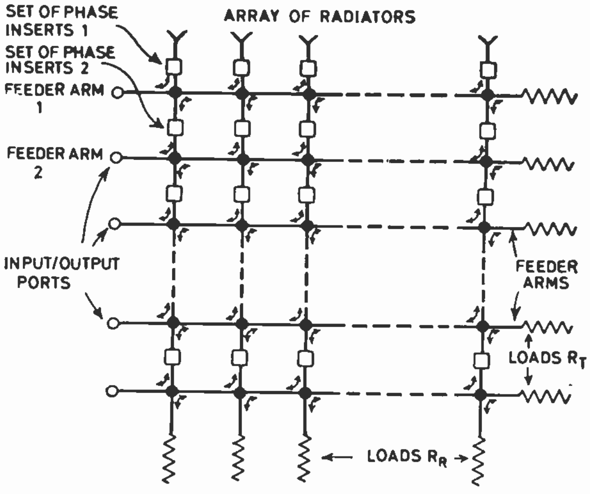


Fig. A. Schematic of Maxson multi-beam antenna.

should not remain independent. The required phases and amplitudes of the constituent beams would be obtained by a suitable network on the input/output terminal ports. Since each constituent beam is independent, impedance matching can be handled independently at the appropriate port.

We now have to consider if the array should be supergain, that is, be fully coupled to space and have the gain or receiving area proper to the beamwidth, or if it should be only superdirective, that is, be weakly coupled to space and have a low gain or receiving area, certainly no higher than that normally associated with the actual aperture. (If amplifiers are used on the elements, or on the input/output ports of the matrix, these questions must be redefined in terms of signal/internal-noise ratio on reception, or total power efficiency on transmission.) In forming a supergain or superdirective beam two areas of difficulty can be distinguished. One area is that of the tolerances and frequency dependence of the required distribution[2] for the superdirective but not supergain antenna; these are the tolerances and frequency functions of a system of signal analysis, which need not postulate a particular r.f. transmit/receive system at all. The other area is that of the impedance seen by the antenna in so far as it is made capable of coupling power into the spherical wave modes required to form the beam in space, irrespective of its being actually designed to transmit. The fundamental difficulty of the supergain antenna lies in the impedances of these modes when they exist on the abnormally small surface surrounding the antenna[3], (e.g. the high Q of a very short antenna radiating only the fundamental dipole mode). Both areas of difficulty have been shown to become acute as the degree of superdirectivity or supergain is increased, and prohibitively acute as the size of the aperture in wavelengths is increased. No one so far as we know has yet made a superdirective antenna of even 2λ aperture; however, we agree with Mr. Redgment† in thinking that the problem is still worth considering, possibly for small receiving antennas for applications in which the system would be externally noise-limited.

† 'Superdirective arrays', *The Radio and Electronic Engineer*, **36**, No. 1, pp. 33–36, July 1968. (Letter.)

Returning to the Maxson array, if we design the feeder arms for all beams such that all the power is coupled through to the aperture, we would potentially have a supergain antenna, with its attendant very high Q which we expect to appear in the matching network at the terminal ports of the wide angle and imaginary beams. In talking of an imaginary beam we are of course merely talking of a lobe pattern in real space which can be formally represented as the side-lobes of a beam in imaginary space. The expected high Q is due to the aperture impedance for radiation from aperture distributions giving these 'imaginary beam side-lobe' patterns. We can, if we wish, damp the matching networks of each of the beam terminal ports appropriately and independently and so control bandwidth/efficiency ratio. Another way of damping the terminal port of an 'imaginary beam side-lobe' pattern is to design the network coupling factors so that the port is coupled only partially to the aperture and the remaining coupling is to the end load. We could formally represent this as coupling the imaginary space power fraction into the end load and only the 'real space side-lobe' power fraction into the aperture. This way of thinking leads to a particular 'law' of damping.

The main advantage of looking at a superdirective antenna in this way, as opposed to considering a single phase and amplitude distribution across the aperture, is that this network enables one to control the impedance matching, phase, and amplitude requirements for the spatial components of the pattern independently. This seems a better approach than trying to control the elements of the array independently, since the spatial components of the pattern are independent, and the elements of the array most certainly are not.

The aperture distribution corresponding to any symmetric pair ($\pm \sin\theta$) of $(\sin\phi)/\phi$ patterns is a sine wave of order which increases with $\sin\theta$. The sine waves corresponding to the 'imaginary beam side-lobe' pairs are those whose complete period is less than a wavelength of aperture. Hence the analysis can be made alternatively in terms of the modes of the aperture and the cause of the high Q seen in another form.

The views expressed above are those of the writers and should not be taken as representing the policy of the Ministry of Defence.

J. CRONEY, PH.D.
E. A. KILLICK, B.SC.
H. SALT, B.SC.

Admiralty Surface Weapons Establishment,
Funtington, Chichester, Sussex.

19th September 1968.

References

1. Allen, J. L., 'A theoretical limitation on the formation of lossless multiple beams in linear arrays', *Trans. I.R.E. on Antennas and Propagation* **AP-9** No. 4, p. 350, July 1961.
2. Woodward, P. M. and Lawson, J., 'The theoretical precision with which an arbitrary radiation pattern may be obtained from a source of finite size', *J. Instn Elect. Engrs*, **95**, Part III p. 363, 1948.
3. Chu, L. J., 'Physical limitations of omni-directional antennas', *J. Appl. Phys.*, **19**, p. 1163, 1948.

STANDARD FREQUENCY TRANSMISSIONS

(Communication from the National Physical Laboratory)

Deviations, in parts in 10^{10}, from nominal frequency for **September 1968**

September 1968	24-hour mean centred on 0300 U.T.			September 1968	24-hour mean centred on 0300 U.T.		
	GBR 16 kHz	MSF 60 kHz	Droitwich 200 kHz		GBR 16 kHz	MSF 60 kHz	Droitwich 200 kHz
1	− 300·0	0	− 0·2	16	− 300·1	− 0·2	—
2	− 300·0	− 0·1	− 0·2	17	− 300·1	0	0
3	− 300·0	0	− 0·2	18	− 300·0	− 0·1	− 0·1
4	− 300·0	0	− 0·3	19	− 300·1	− 0·1	− 0·2
5	− 300·0	0	− 0·2	20	− 300·0	− 0·1	− 0·2
6	− 299·9	0	− 0·1	21	− 300·0	0	− 0·2
7	− 300·0	+ 0·1	0	22	− 300·1	− 0·1	− 0·2
8	− 300·0	0	− 0·1	23	− 300·2	0	− 0·2
9	− 299·9	0	—	24	− 300·1	0	− 0·2
10	− 300·0	0	—	25	− 300·0	0	− 0·2
11	− 299·9	0	− 0·1	26	− 300·0	− 0·1	− 0·2
12	− 300·1	0	− 0·1	27	− 299·9	0	− 0·2
13	− 300·0	0	− 0·1	28	− 300·1	0	− 0·2
14	− 300·0	0	—	29	− 300·0	0	− 0·2
15	− 300·1	0	—	30	− 300·1	0	− 0·2

Nominal frequency corresponds to a value of 9 192 631 770·0 Hz for the caesium F,m (4,0)–F,m (3,0) transition at zero field.

U.D.C. 621.372.57

The Design of Active Filters using the In-line Pole Approximation

By

L. N. HULLEY,
B.Sc., M.Sc., C.Eng., M.I.E.E.†

AND

G. MUSGRAVE,
B.Tech., M.Sc.†

Summary: A synthesis is presented which has enabled a practical switchable filter to be realized which will give an optimum linear phase response together with satisfactory amplitude and transient response using a computed in-line pole distribution. Theoretical and practical characteristics are compared for a fourth- and sixth-order filter and the appropriate circuit diagrams are given. The application of this type of filter is discussed with particular reference to the field of medical electronics.

List of Symbols

K	constant
σ	abscissa of pole locations on complex-frequency plane
ω	ordinate of pole locations on complex-frequency and real frequency
ω_c	cut-off frequency (−3 dB)
Am_4	amplitude response of fourth-order filter
Am_6	amplitude response of sixth-order filter
Ph_4	phase response of fourth-order filter
Ph_6	phase response of sixth-order filter
$Tran_4$	step response of fourth-order filter
$Tran_6$	step response of sixth-order filter
A	open-loop gain of amplifier
A_1	amplifier gain when feedback network is ineffective but connected
β	feedback fraction
$F(s)$	polynomial in (s)
s	complex operator
P	pole locations
ϕ	phase response

1. Introduction

The synthesis of passive networks has progressed considerably from the early work of Butterworth[1] and Darlington[2] to the state where several writers[3-5] have published tabulated normalized values of circuit components for the realization of mathematically based approximations. Szentirmai[6] and Saraga[7] have outlined accurate methods of component evaluation using computational techniques; however, for low frequencies the resulting values for denormalized inductance are prohibitive. On account of the comparatively large physical size of wound inductors difficulty is also encountered when inductance is required in integrated circuits. A gyrator and a capacitor are often adopted to simulate the required value of inductance as they can be more readily accommodated on a silicon chip. This has led to the development of active R–C networks and in consequence has resulted in some cataloguing of transfer functions and associated circuits.[8,9] Transfer functions have also been realized using negative impedance convertors.[10,11]

Generally, filter design has evolved in the field of telecommunications where phase response is relatively unimportant, as the human ear is insensitive to phase difference of speech waveforms. But when the display of the waveform is a requirement, it is desirable to have as little phase change with frequency and delay of transient phenomena, as possible. In order to achieve this ideal the filter characteristic must be such that it gives a linear-phase constant-time delay response. One field in which display is an important requirement is medical electronics, where cardiac and neurological waveforms are invaluable in the diagnosis of malfunctions and have become a necessity in major surgery. The pursuit of these requirements has resulted in the use of 'maximally-flat delay',[12] 'in-line pole'[13] and 'ultraspherical approximations'[14] in filter synthesis. Also in the field of medical electronics there are many applications where level-sensing devices are triggered by transient phenomena of the network. Since the actual firing level of such apparatus as cardiac rate-meters etc., is arbitrarily decided by the operator, a very rapid decay of any ringing which occurs is necessary in order to avoid multiple triggering of the sensing device. This is an added requirement to the normal display, constant time delay criteria.

From the work of Holt,[15] who gives results of maximally-flat delay and in-line pole approximations,

† Postgraduate School of Studies in Electrical and Electronic Engineering, University of Bradford.

the latter may be concluded to be marginally better, and in the light of the display and ringing requirements a method of synthesis using an in-line pole approximation (Fig. 1) will be described. The circuit configuration will be such as to facilitate selection of cut-off frequencies with identical characteristics. A practical circuit design for a switchable in-line pole low-pass filter is given, and the experimental results are compared with computed characteristics.

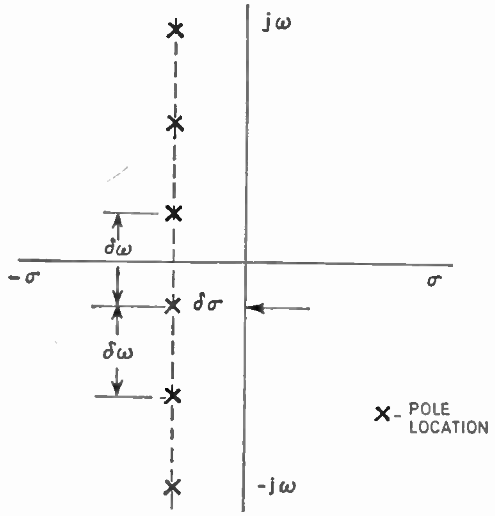


Fig. 1. The in-line pole distribution.

2. The In-line Pole Array

The derivation of the in-line pole approximation is analogous with the field associated with a parallel plate capacitor in concept, in the limit of an infinite array, and consequently some flexibility in pole location is available. This feature differs from the other approximations used in insertion loss theory (such as the Butterworth, Chebyschev, Bessel and ultraspherical approximations) which are derived from mathematical relationships and, with the exception of ultraspherical, polynomials are inflexible in that relative pole location is fixed. Guilleman[13] presents the in-line pole distribution as an alternative linear phase approximant to the Bessel function and mathematically rationalizes the ripple factor for phase and amplitude for discrete spectra. By considering the appropriate phasors on the complex-frequency plane the expression for the amplitude response of a discrete spectrum is given as

$$f^2(\omega) = \frac{K}{\Pi[\sigma^2+(\omega-\omega_c)^2]} \qquad \ldots\ldots(1)$$

and the phase response as

$$\phi = \sum \tan^{-1}\frac{\omega-\omega_c}{\sigma} \qquad \ldots\ldots(2)$$

This has been numerically evaluated by means of a digital computer for $n = 4$ and 6 where the amplitude response is a function of σ and $\delta\omega$. Generality is preserved if $\delta\omega$ is normalized to unity and σ is considered as the variable. Thus the effect on conjugate pairs is to preserve the natural frequency but alter the selectivity. The resultant normalized amplitude and phase responses are given for $n = 4$ and $n = 6$ as:

$$\text{Am}_4 = 10\log\frac{(\sigma^2+1)^2(\sigma^2+9)^2}{[\sigma^2+(\omega-1)^2][\sigma^2+(\omega+1)^2][\sigma^2+(\omega-3)^2][\sigma^2+(\omega+3)^2]} = 10\log\frac{Y_1}{Y_2} \qquad \ldots\ldots(3)$$

$$\text{Am}_6 = 10\log\frac{Y_1(\sigma^2+25)^2}{Y_2[\sigma^2+(\omega-5)^2][\sigma^2+(\omega+5)^2]} \qquad \ldots\ldots(4)$$

$$\text{Ph}_4 = -\frac{180}{\pi}\left[\tan\left(\frac{\omega-1}{\sigma}\right)+\tan\left(\frac{\omega+1}{\sigma}\right)+\tan\left(\frac{\omega-3}{\sigma}\right)+\tan\left(\frac{\omega+3}{\sigma}\right)\right] \qquad \ldots\ldots(5)$$

$$= -\frac{180}{\pi}A$$

$$\text{Ph}_6 = -\frac{180}{\pi}A+\tan\left(\frac{\omega-5}{\sigma}\right)+\tan\left(\frac{\omega+5}{\sigma}\right) \qquad \ldots\ldots(6)$$

The evaluated results are shown in Figs. 2, 3, 4 and 5 respectively.

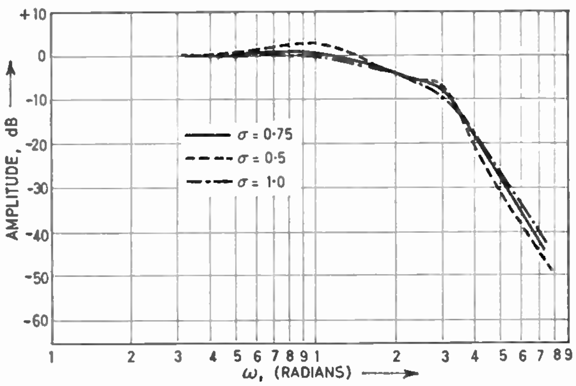


Fig. 2. $N = 4$ In-line pole amplitude response.

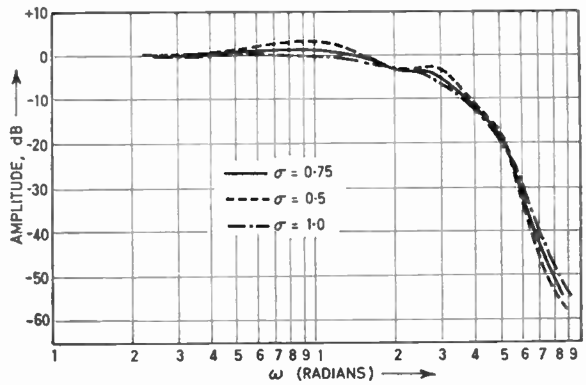


Fig. 3. $N = 6$ In-line pole amplitude response.

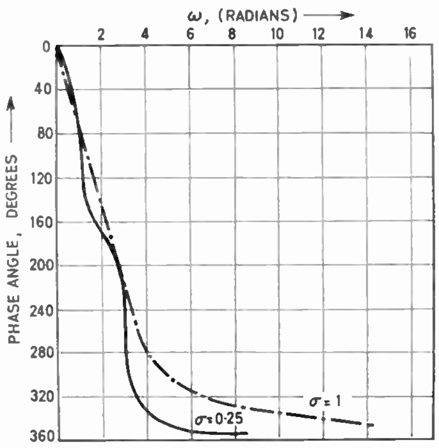


Fig. 4. $N = 4$ In-line pole phase response.

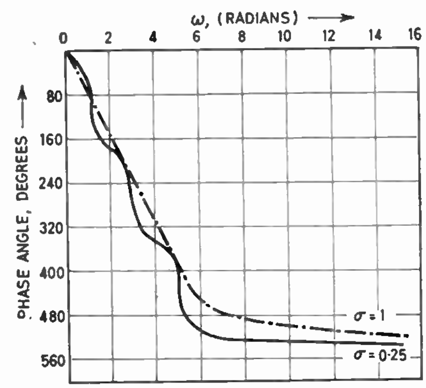


Fig. 5. $N = 6$ In-line pole phase response.

2.1 *Consideration of Computational Results*

Holt[15] considered a filter with the array location such that $\sigma = 1$ but with a pole distribution arranged so as to obtain the -3dB point at $\omega = 1$, thus enabling comparison to be made with conventional filters, whilst Lerner[16] in the design of band-pass filters briefly investigates the displacement of poles from the imaginary axis, namely, $\sigma = 0{\cdot}75$, 1 and 1·25. The latter show a high ripple content which is due to the frequency distribution of the poles resulting from the consideration of bandwidth criteria. Figures 6 and 7 give the normalized amplitude responses for the in-line arrays at representative locations from the $j\omega$-axis (namely $\sigma = 0{\cdot}25$, 1·0 and 2·0). It can be seen that the best rejection occurs for $\sigma = 0{\cdot}25$ in the stop band but, nevertheless, the inherent ripple at the band edges in this proximity of the imaginary axis reduces the rejection immediately outside the pass-band. The $\sigma = 1$ location gives only 5 dB less rejection at the asymptote and better overall discrimination as a result of reduced ripple content.

The phase response was calculated by using eqn. (4) for the same range of array locations, and the results are shown in Figs. 4 and 5. Using Guilleman's theoretical ripple factor the resultant calculations are given in Table 1.

Table 1

Array location σ	Ripple factor
0·25	± 45%
0·5	± 20%
1·0	± 8·0%
1·5	± 4·5%

It can be seen from the graphs (Figs. 4, 5) that the phase is almost linear for $\sigma = 1$ which confirms the choice of array locations for the criteria described for medical electronics.

The transient response for the same set of in-line poles was also evaluated using a digital computer,

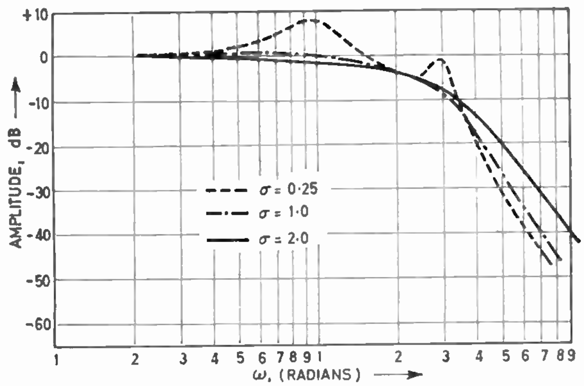


Fig. 6. $N = 4$ In-line pole amplitude response.

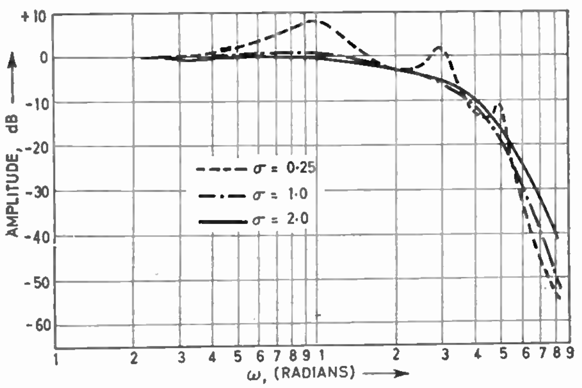


Fig. 7. $N = 6$ In-line pole amplitude response.

the step response functions in the time domain being given by

$$\text{Tran}_4 = (1 - 0{\cdot}125e^{-\sigma t})\Big[(\sigma^2+9)\{\cos(t) + \sigma\sin(t)\} - (\sigma^2+1)\Big\{\cos(3t) + \frac{\sigma}{3}\sin(3t)\Big\}\Big] \quad \text{......(7)}$$

$$= (1 - 0{\cdot}125\, e^{-\sigma t})(A - B)$$

$$\text{Tran}_6 = (1 - e^{-\sigma t})\Big[\frac{(A(\sigma^2+25)}{192} - \frac{B(\sigma^2+25)}{128} + \frac{(\sigma^2+1)(\sigma^2+9)}{384}\Big\{\cos(5t) + \frac{\sigma}{5}\sin(5t)\Big\}\Big] \quad \text{......(8)}$$

The resultant responses are given in Figs. 8 and 9 from which it can be seen that the amplitude of the ringing is as expected, inversely related to σ provided ω is constant and equal to unity. Also, on account of the interaction of pole-pairs the peak of the ringing is depressed tending towards a rectilinear waveform. It can be seen by comparing Figs. 8 and 9 for the same value of σ that the higher the order of the system the nearer a square-wave of decaying envelope is obtained.

Here again the location of the array at unit distance from the imaginary axis produces a satisfactory response in that the amplitude of the second cycle is exceedingly small. There is thus little chance of error occurring, due to sporadic firing of level-sensitive circuits in rate-recording medical equipment, if this approximation is used for any associated filters.

In the light of the foregoing results and considerations, it was decided that the optimum location of an in-line array to meet the requirements of a 'biological filter' occurs at $\sigma = 1$. It was therefore decided to adopt the in-line pole array with this location in the subsequent filter realizations.

3. Active Realization of the In-line Pole Array

In trying to realize the higher-order filter characteristics the practical restrictions on the designs have been extremely difficult to overcome, particularly if a switchable filter is required. The difficulties have varied from close tolerance capacitors of inconvenient value to uneconomically large numbers of capacitors.[16]

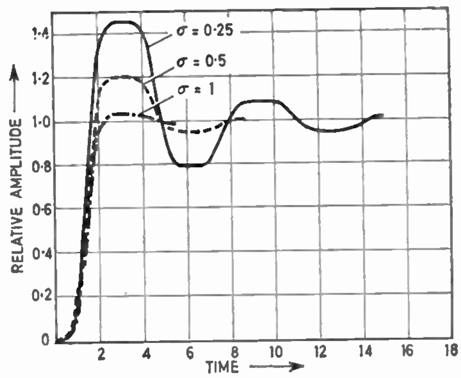


Fig. 8. $N = 4$ In-line pole step response.
Step height = 1·0.

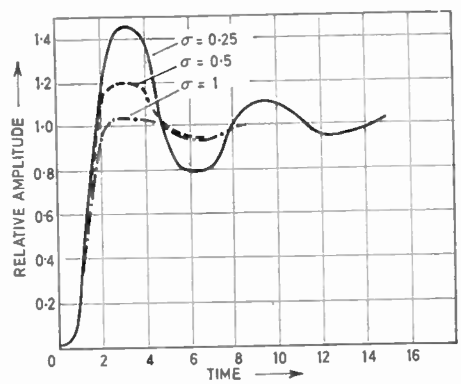
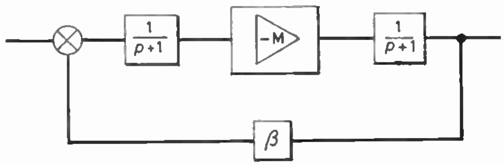
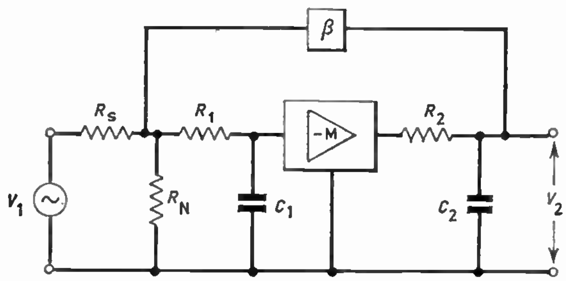


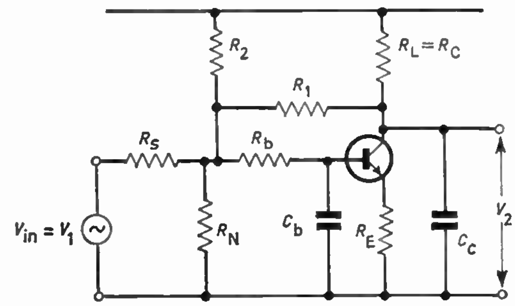
Fig. 9. $N = 6$ In-line pole step response.
Step height = 1·0.



(a) Basic configuration of a system for generating conjugate pole-pairs (the pole-forming networks have been normalized).



(b) Schematic diagram giving the de-normalized arrangement for generating conjugate pole-pairs.



(c) Circuit diagram showing the connection details of a transistor for the generation of conjugate pole-pairs.

Fig. 10.

In the field of medical electronics and indeed in other specialized instrumentation fields there is a requirement for switchable filters. But the designs of the insertion-loss filter characteristics require a large number of capacitors for a moderate range of cut-off frequencies. Even in the telecommunications field where the characteristics are fixed, the available designs specify non-standard capacitors which cause inconvenience particularly for small quantity production. In such a case it would be preferable to be able to select the required capacitors from batches of devices manufactured with nominal values.

3.1 *Choice of Circuit*

To produce the correct time-constants for the pole-forming networks and to still use preferred values of capacitors the design philosophy must give sufficient freedom for variation of the resistors. Although Sallen and Key[8] established the basic layout of the pole-forming networks before the amplifier and Smith[19] used this to advantage in a switchable application, the network configuration shown in Fig. 10(a) minimizes the number of capacitors which dictate the approximation. This design technique makes a network which generates an in-line pole distribution practicable.

The realization of rational coefficients using feedback amplifiers can be broadly categorized into the following three main groups according to the amplifier gain:

(1) amplifiers with low-valued gains which are accurately defined and with zero phase shift;[17]

(2) amplifiers with very high gains which are intended to simulate a theoretical ideal of infinite gain;[18] and

(3) amplifiers with intermediate values of negative gain[19] and with a phase shift of π radians.

At first sight it would appear that amplifiers with low positive gains offer the best solution, particularly as a gain of only $+2$ is required to generate all the families of approximations.[18] There is however with such a system a distinct possibility of instability arising when switches and different capacitors have to be included in the circuit, since the sensitivity to component and gain variation is considerable.

An attractive approach using amplifiers of the second category has been given by Holt[18] who has synthesized fourth- and sixth-order low-pass functions by the use of a single high-gain amplifier. Multiple feedback paths are used, but it is found that the values of the capacitive elements in the pole-forming networks are awkward. The internal poles of the system could well be very near instability and errors in component values and strays arising from the switching facility could well give rise to oscillations.

It is possible, by using a reasonable negative gain, to generate all but the most selective second-order pole-pairs by using only two capacitors. These are inserted in the forward path of an amplifier as shown in Fig. 10(b).

It is impossible to synthesize a transfer function with rational terms in only the denominator if capacitor pole-forming networks are only to be used in the feedback loop. This can be seen by considering the feedback loop, to be complex, while the forward gain A may be real positive or negative.

Then the overall gain is

$$A = \frac{A_1}{1-\beta A_2}$$

where $A_1 \neq A_2$ because of interaction between the feedback loop and amplifier circuit. To be consistent with the above, the numerator of A_1 must be constant, but the numerator of A_2 may be a function of s, where s is the complex variable. They must however both possess an identical function in s in both their denominators. Conversely if β is complex and A_2 is real then the function of β may take the following forms:

$$\beta_1 = \frac{1}{f(s_1)} \qquad \beta_2 = \frac{f(s_2)}{f(s_3)} \qquad \beta_3 = f(s_3$$

(For this latter identity gyrators or inductors would be required.) The latter may be dismissed as it implies the use of inductance which is contrary to the objective of using active circuits. Both β_1 and β_2 will yield expressions for A which contain terms in s, both in numerator and denominator; it is only by factor cancellation that these terms will disappear. It can be concluded, therefore, that the best method of generating low-pass functions in cases where reactive components have to be minimized is that of making A_2 of the form $K/f(s)$. A transfer function of this type may be realized by the basic R–C circuit configuration which will give a time-constant $T = RC$. Analysis of the circuit shown in Fig. 10(b) gives a transfer function

$$F(A') = \frac{K\left[\dfrac{1}{T_1 s+1}\right]\left[\dfrac{1}{T_2 s+1}\right]}{1+\left[\dfrac{1}{T_1 s+1}\right]\left[\dfrac{1}{T_2 s+1}\right]\beta . A} \qquad \text{......(9)}$$

where $\dfrac{1}{Ts+1}$

is a capacitor-resistor time-constant circuit given by:

$$\frac{K/T_1 T_2}{s^2+\left(\dfrac{1}{T_1}+\dfrac{1}{T_2}\right)s+\dfrac{1+\beta A}{T_1 T_2}} \qquad \text{......(10)}$$

and from the Appendix

$$P = -\tfrac{1}{2}\frac{1}{T_1 T_2} \pm \tfrac{1}{2}\frac{1}{T_1 T_2}[(T_1-T_2)^2 - 4T_1 T_2(1+Kg_{21})]^{\frac{1}{2}} \qquad \text{......(11)}$$

Thus by adjusting values of T_1, T_2 and βA any pole-pair location may be established. No generality is lost by letting $T_1 = T_2$ and in fact this condition gives the minimum active-gain requirement as

$$F(A) = \frac{K/T^2}{s^2+\dfrac{2}{T}s+\dfrac{1+\beta A}{T^2}} \qquad \text{......(12)}$$

$$P = \frac{1}{T} \pm \frac{1}{T}\mathrm{j}(\beta A)^{\frac{1}{2}} \qquad \text{......(13)}$$

It can be seen that the abscissa is described by the values of R and C and the normalized value of the ordinate by $F-1$ where F is the feedback factor of the amplifier.

It can also be seen at this stage that the pole locations for a constant-T coincides with the in-line pole array for the same value of T. The worthwhile possibility arises therefore of synthesizing a complete multiple-order filter with one value of T and possibly one value of C. (This idea could lend itself to integrated circuit techniques.) It would thus be possible to choose a range of cut-off frequencies from simple preferred values of paper capacitors.

A transistor connected in a common-emitter configuration with an additional resistor inserted in the emitter circuit can generate an accurate voltage gain if driven from a suitable low-impedance source. The bulk resistance of the emitter and thus r_e whilst being a function of emitter current is consistent for a given transistor type. Further, any variations can be reduced in fact by operating the transistor at as high an emitter current as possible. The resistance r_e only accounts for a relatively small fraction of the total resistance in the emitter circuit, R_E' which is equal to $r_e + R_E$. The total emitter resistance, R_E', can then be arranged so as to raise the input resistance of the transistor in order that a pole-forming network may be included before the base without insertion loss occurring. The basic circuit is given in Fig. 10(c) and the detailed associated voltage transfer characteristic is developed in the Appendix and is

$$F(s) = \frac{g_{21}\dfrac{R_1//R_n}{R_s+R_n//R_1}}{T_b T_c s^2+(T_c+T_b)s+1+g_{21}\dfrac{R_s//R_n}{R_1+R_s//R_n}} \qquad \text{......(14)}$$

which results in a pole-pair location

$$P = -\tfrac{1}{2}\left(\frac{1}{T_b}+\frac{1}{T_c}\right) \pm \tfrac{1}{2}\left[\left(\frac{1}{T_b}+\frac{1}{T_c}\right)^2 - \frac{4}{T_b T_c}\left(1+g_{21}\frac{R_s//R_n}{R_1+R_s//R_n}\right)\right]^{\frac{1}{2}}$$

$$= -\tfrac{1}{2}\left(\frac{1}{T_b}+\frac{1}{T_c}\right) \pm \tfrac{1}{2}\frac{1}{T_b T_c}\left[T_b^2+T_c^2+ +2T_b T_c - 4T_b T_c g_{21}\left(\frac{R_s//R_n}{R_1+R_s//R_n}\right)\right]^{\frac{1}{2}} \qquad \text{......(15)}$$

This may be arranged to give an expression for P as follows:

$$-\frac{1}{2}\left(\frac{1}{T_b}+\frac{1}{T_c}\right)\pm\frac{1}{2}\frac{1}{T_bT_c}\left[(T_b+T_c)^2 - 4T_bT_c\left(1+g_{21}\frac{R_s//R_n}{R+R_s//R_n}\right)\right]^{\frac{1}{2}}$$

Expanding the squared parenthesis term we have

$$\left[(T_b-T_c)^2-4g_{21}\frac{R_sR_n}{R_1+R_s//R_n}T_bT_c\right]^{\frac{1}{2}} \qquad \text{......(16)}$$

If the time-constants are equal then the first term is zero and any gain greater than zero will cause vertical migration of the poles thus increasing selectivity.

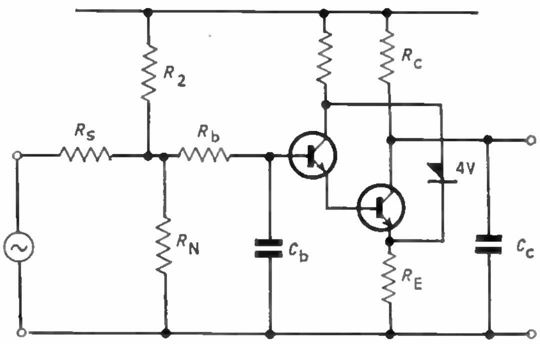


Fig. 11. Circuit diagram showing two active devices connected as a Darlington pair to generate a higher input impedance and higher voltage gain.

3.2 *Pole Sensitivity to Component Variation*

The pole migration behaviour has been investigated by means of evaluating the root-locus equations for variation of A and tolerance variation of either component of each and both pole forming networks. The conclusions reached were that regardless of magnitude of component tolerance deviation from design centre or of variation of gain, the system would remain stable and would be insensitive to a component variation as compared with systems where amplifiers with positive gains are used.

4. Practical Results

Taking a sixth-order in-line pole array $\sigma = 1$ array location will give a normalized cut-off frequency (-3dB) of almost two radians per second. The pole locations for this array will be

$$1 \pm j5 \quad ; \quad 1 \pm j3 \quad ; \quad 1 \pm j1$$

Hence

$$(\beta A)^{\frac{1}{2}} = \quad 5 \quad ; \quad 3 \quad ; \quad 1$$

$$\beta A \;= \quad 25 \quad ; \quad 9 \quad ; \quad 1$$

The two lower gains can be realized accurately by single transistor stages but the higher βA product has to be realized by a compound transistor connection.[22-22] A boot-strapped common-collector configuration is shown in Fig. 11. It can be shown that a high voltage-gain can be achieved as R_E' can be reduced. Also the input resistance for that reduced value of R_E' is much greater than for a single transistor. This provides more flexibility in the choice of R_E and thus R_B and R_L may be increased. A circuit diagram using the principles given here is shown in Fig. 12. It can be seen that a flat response and four

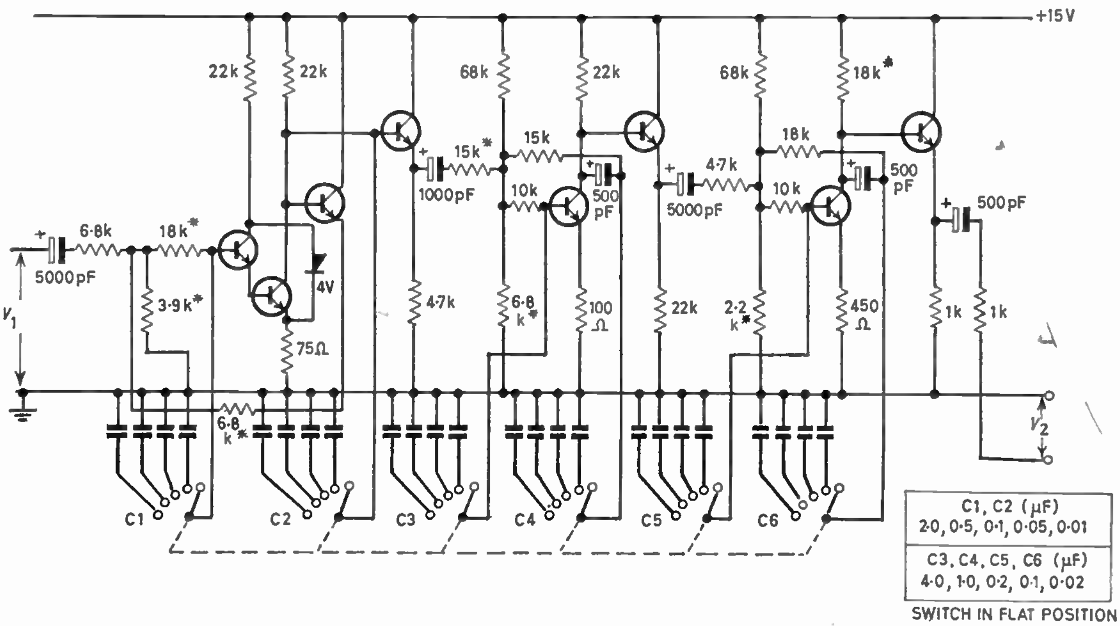


Fig. 12. Circuit diagram of the $N = 6$ in-line pole filter.
* shows high-stability type resistors.

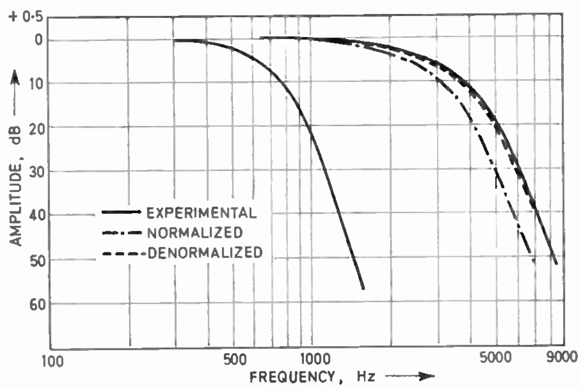


Fig. 13. Curves of $N = 6$, $\delta = 1$ in-line pole distribution.

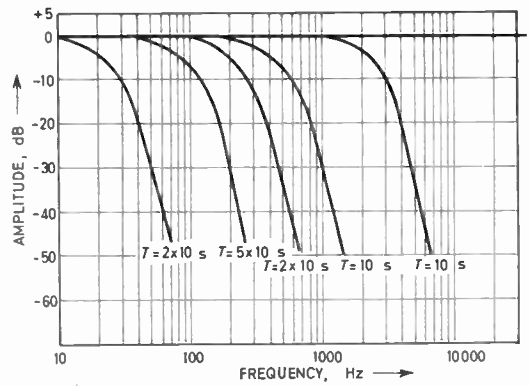
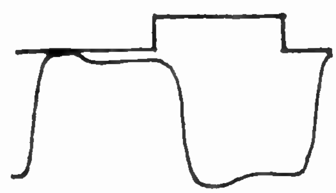
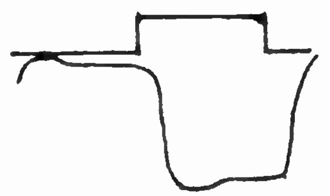


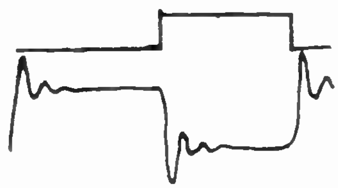
Fig. 14. Curves of $N = 6$, $\delta = 1$ in-line pole filter.



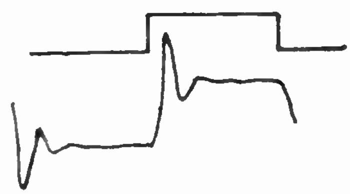
(a) H.t. = 14 V.



(b) H.t. = 20 V.



(c) After stage 2.



(d) After stage 1.

Fig. 15. In-line pole step response.

cut-off frequencies have been readily provided with convenient value of capacitance. A comparison of the theoretical and actual amplitude characteristics are given in Fig. 13 and the range of characteristics available by switching are given in Fig. 14. The square-wave responses are presented in Fig. 15. From these it can generally be concluded that the method of synthesis and of practical realizations are both satisfactory.

5. Conclusions

The design technique presented has allowed a range of active R–C filters with useful characteristics to be conveniently realized. It has been found possible to use standard values of capacitors in these filters and less components are needed than in many other methods of synthesis which have been described in the literature. Satisfactory agreement of theoretical and practical results has been achieved.

The approximation and design philosophy which have been adopted have enabled the criteria of constant time delay and acceptable responses to step and impulse signals for filters in the field of medical electronics to be met. Also they have been achieved with an economical design which can well be utilized for switchable filter requirements in other fields and indeed where another approximation may be deemed more suitable.

6. Acknowledgments

The authors would like to thank Professors D. P. Howson and G. N. Patchett for their permission to use the facilities of the School of Electrical and Electronic Engineering of the University of Bradford.

7. References

1. Butterworth, S., 'On the theory of filter amplifiers', *Exptl. Wireless and Wireless Engng*, **7**, p. 536, October 1930.
2. Darlington, S., 'Synthesis of reactance 4-poles which produce prescribed insertion loss characteristics including special applications to filter design', *J. Maths Phys.*, **18**, p. 275, September 1939.
3. Weinberg, L., 'Network Analysis and Synthesis', (McGraw-Hill, New York, 1962).
4. Skwirzynski, J. K., 'Design Theory and Data for Electrical Filters', (Van Nostrand, New York, 1955).
5. Geffe, P. R., 'Simplified Modern Filter Design', (Iliffe, London, 1964).
6. Szentirmai, G., 'New developments in the theory of filter synthesis', Conference on Electrical Networks, University of Newcastle-upon-Tyne, September 1966.
7. Saraga, W., 'A new approach to insertion loss ladder filter design', *ibid.*
8. Sallen, R., and Key, E. L., 'A practical method of designing RC active filters', *Trans. Inst. Radio Engrs on Circuit Theory*, **CT-11**, p. 74, March 1964.
9. Foster, E. J., 'Active low pass filter design', *Trans. Inst. Elect. Electronics Engrs on Audio*, **AU-13**, No. 5, p. 104, September–October 1965.
10. Yanagisawa, T., 'RC active networks using current inversion type negative impedance convertor', *Trans. I.R.E.*, **CT-4**, p.140, 1957.
11. Linvill, J. G., 'RC active filters', *Proc. I.R.E.*, **42**, p. 555, July 1954.
12. Storch, L., 'Synthesis of constant-time-delay ladder networks using Bessel polynomials', *Proc. I.R.E.*, **42**, p. 1666, November 1954.
13. Guilleman, E. A., 'Synthesis of Passive Networks', Chapter 14, p. 639 (Wiley, New York, 1965).
14. Johnson, D. E. and Johnson, J. R., 'Low pass filters using ultraspherical polynomials', *Trans. I.E.E.E.*, **CT-13**, p. 364, December 1966.
15. Holt, A. G. J., 'A comparison of five methods of low-pass filter design', *The Radio and Electronic Engineer*, **27**, p. 167, March 1964.
16. Lerner, R. M., 'Band pass filters with linear phase', *Proc. I.E.E.E.*, **52**, p. 249, March 1964.
17. Hakim, S. S., 'RC active filters using an amplifier as the active element', *Proc. I.E.E.*, **112**, p. 901, May 1965.
18. Paul, R. J. A., 'Active network synthesis using one port RC networks', *Proc. I.E.E.*, **113**, p. 83, January 1966.
19. Holt, A. G. J., 'RC active networks', Conference on Electrical Networks, University of Newcastle-upon-Tyne, September 1966.
20. Smith, K. D., 'An active Bessel-Butterworth filter', I.E.E. International Convention Record, March 1964.
21. Pearlman, R. A., 'Some properties and circuit applications of super-alpha composite transistors', *Trans. I.R.E. on Electron Devices*, **ED-2**, p. 25, January 1953.
22. Gandhi, S. K., 'Darlington compound connection for transistors', *Trans. I.R.E.*, **CT-4**, p. 291, September 1957.
23. Yu, P. K. and Suen, C. Y., 'Analysis of Darlington configuration', *Electronic Engng*, **40**, p. 38, January 1968.

8. Appendix

From Fig. 7(b) the overall gain may be written as

$$A = \frac{\dfrac{g_{21}}{(T_c s+1)(T_b s+1)}\,\dfrac{R_1//R_n}{R_s+R_1//R_n}}{1+\dfrac{g_{21}}{(T_c s+1)(T_b s+1)}\,\dfrac{R_s//R_n}{R_1+R_s//R_n}}$$

where T_b and T_c are the time-constants of the basic circuit shown in Fig. 7(c), suffixes 'b' and 'c' referring to base and collector circuit.

Rewriting we have

$$A = \frac{g_{21}\,\dfrac{R_1//R_n}{R_s+R_n//R_1}}{T_b T_c s^2+(T_c+T_b)s+1+g_{21}\,\dfrac{R_s//R_n}{R_1+R_s//R_n}}$$

Thus the transfer function may be written as:

$$F(p) = \frac{g_{21}\,\dfrac{R_1//R_n}{R_s+R_n//R_1}\,\dfrac{1}{T_b T_c}}{s^2+\left(\dfrac{1}{T_b}+\dfrac{1}{T_c}\right)s+\dfrac{1}{T_b T_c}\left(1+g_{21}\,\dfrac{R_s//R_n}{R_1+R_s//R_n}\right)}$$

The denominator gives the pole locations as:

$$P = -\tfrac{1}{2}\left(\frac{1}{T_b}+\frac{1}{T_c}\right)\pm\tfrac{1}{2}\left[\left(\frac{1}{T_b}+\frac{1}{T_c}\right)^2 - \frac{4}{T_b T_c}\left(1+g_{21}\,\frac{R_s//R_n}{R_1+R_s//R_n}\right)\right]^{\frac{1}{2}}$$

$$= -\tfrac{1}{2}\left(\frac{1}{T_b}+\frac{1}{T_c}\right)\pm\tfrac{1}{2}\,\frac{1}{T_b T_c}\left[T_b^2+T_c^2+2T_b T_c - 4T_b T_c - 4T_b T_c g_{21}\left(\frac{R_s//R_n}{R_1+R_s//R_n}\right)\right]^{\frac{1}{2}}$$

For minimum gain criteria

$T_b = T_c = T.$

Therefore

$$P = -\frac{1}{T}\pm\tfrac{1}{2}\,\frac{1}{T^2}\left[-g_{21}4T^2\left(\frac{R_s//R_n}{R_1+R_s//R_n}\right)\right]^{\frac{1}{2}}$$

$$= -\frac{1}{T}\pm \mathrm{j}\,\frac{1}{T^2}\left[g_{21}4T^2\left(\frac{R_s//R_n}{R_1+R_s//R_n}\right)\right]^{\frac{1}{2}}$$

Manuscript first received by the Institution on 2nd May 1968 and in final form on 22nd July 1968. (Paper No. 1213/CC22).

Television and Education

In May 1952 the Institution's *Journal* welcomed the first limited television programme for schools, sent out by the B.B.C. on a special wavelength with sound on land line to six secondary schools in Middlesex. This was the beginning of educational television in this country and it was followed in 1957 by broadcast programmes for schools by the B.B.C. and Independent Television. These services have greatly expanded over the past ten years: programmes cover a wide range of subjects and all age groups.

Television for schools is thus a recognized aid for the teacher but the broadcast programmes, even those regionally radiated, have certain disadvantages for some purposes. Because primary and secondary education is controlled on a 'local authority' basis throughout Great Britain, several authorities have formed their own educational television (E.T.V.) services.

The first large E.T.V. network was opened in Glasgow last year and smaller schemes are operating on an experimental basis elsewhere. The most ambitious system in Britain so far, however, has just been inaugurated by the Inner London Education Authority and ultimately a network of coaxial cables will link all the 1,300 I.L.E.A. primary and secondary schools, colleges of further and higher education, adult education institutes, recreational institutes and youth centres to the E.T.V. Centre.

The first year's transmissions will be seen only by some 300 schools and colleges in three London Boroughs but in September 1969 most of the remaining establishments will be linked up. The system is capable of carrying at least seven channels: two are to be used to relay educational programmes of the B.B.C. and Independent Television (from 1969); one is to be reserved for use by the Universities in London, the polytechnics, colleges of education and other higher education colleges; the remainder are for transmitting programmes made at the E.T.V. Centre. During the first year, programmes will be recorded and transmitted from a temporary centre in Islington, but from September 1969 the service will transmit from a permanent centre at which studios will become operational in 1970.

The E.T.V. Centre has one production studio, a training studio, a master control room which houses the video tape recorders and transmission switcher, a room assigned to the General Post Office housing the modulators and transmission equipment, a maintenance workshop, and a two-camera mobile television unit. Ancillary facilities include a photographic department, graphics and scenic department, rehearsal areas, a film viewing and editing area, and production and administrative offices.

The Production Studio is equipped with three image orthicon E.M.I. type 203 cameras, each having 2 in, 3 in, 5 in, and 8 in lenses. The camera control units are equipped with 'joystick' remote controls which are grouped for operation by one 'vision engineer'. A 16 mm telecine of vidicon type is also remotely controlled by the same engineer. The vision mixer has five inputs and is of the 'A/B' type, with a preview bank. Seven 14 in picture monitors provide viewing of all inputs and give preview and transmission facilities. A comprehensive 'talk-back' system is provided.

A separate sound control room has a 12-input mixer, each input being high or low level, with preset level control, prehear, foldback and echo. The inputs can be routed via a 'red' or 'green' group, and group faders and a main fader allow comprehensive mixing.

The Master Control Room accommodates the 625-line broadcast type synchronizing pulse generator and the distribution amplifiers to send the pulses required to the studio and technical areas. The master mixer is of auto-preview type and can accommodate eight sources; it is normally operated in a 'married' condition but separate picture and sound switching can be employed. There are two four-head broadcast type tape machines (Ampex 1200).

The network has a maximum capacity of nine channels, and is at present capable of the simultaneous transmission of seven channels carrying 625-line monochrome or PAL colour television signals. The channels are in the v.h.f. band of 40 to 140 MHz with a spacing of 10·5 MHz. The vision and sound signals are modulated in an identical manner to B.B.C. 625-line broadcasts.

The incoming cable network to a school is connected to a terminal amplifier where a pilot lamp indicates whether the signals are present or not. From this amplifier the signals are distributed throughout the building to provide a signal level of between 1 mV and 3 mV at each reception point, with an isolation of at least 40 dB between points.

A special E.T.V. receiver has been developed for classroom use. It has a 25 in picture tube, is for 625-line standard only, and incorporates a transistor tuner unit designed to ensure the good adjacent channel rejection which is necessary for the E.T.V. network channel frequencies used. The receiver complies with BS.415 safety requirements for radio and other electronic apparatus. The set has been specified to give good sound and vision response, with improved picture linearity and higher brightness compared with domestic sets and includes a black level clamp.

As continuity of service is of paramount importance, precautions have been taken to ensure that programmes lost for any reason are quickly restored. A simple routine has been devised that will enable the teacher quickly to decide whether a fault is due to the receiver or the network telephone and then the system control centre.

What are the advantages of this kind of E.T.V. scheme over the nationally broadcast schools programmes? Obviously the opportunity for many teachers to participate actively in the conception and production of programmes is very valuable and the I.L.E.A. has 'secondment' arrangements for teachers to produce programmes; courses on production techniques are also held. Many programmes originate (on video tape) in the schools themselves.

From the point of view of the engineering involved, educational television is, in the I.L.E.A. Service, an application of established multi-channel v.h.f. wired television technique—no new 'breakthrough' has been needed, just the use of sound engineering principles in producing an efficient and reliable system. It is, however, an application of television to meeting a particularly worthwhile need.

U.D.C. 621.374:621-52

A Positioning System using Three-valued Codes

By

A. PUGH,

B.Sc., C.Eng., M.I.E.E., M.I.E.R.E.†

Summary: A significant simplification of the computation circuits for positioning systems is achieved by the use of three-valued codes. The use of the odd radix makes it possible to provide a control system symmetrical to positive and negative quantities which relieves the necessity to complement digital information. The paper discusses the construction of suitable cyclic codes and gives the logical design for the necessary conversion logic. Design proposals for the remainder of the control system are also incorporated.

1. Introduction

With digital positioning systems, the principle of computing the error or difference signal is basically very simple. In the case of a mechanical member, its position will be measured by a suitable digital transducer or digitizer. The digitizer output is subtracted from the reference signal to determine the magnitude and direction of the mechanical displacement from the desired or reference position. The resultant error number is used to apply control signals to the driving mechanism in order to bring the position of the member and the reference into coincidence.

However, two practical considerations arise out of this example. Firstly, to eliminate ambiguity in the digitizer, one of the Gray-codes must be used, and secondly, because the digitizer output can be either larger or smaller than the reference, the process of subtraction will be usually accomplished by complementing and adding. In the event of a negative error signal, a further complementing process is necessary to determine the true magnitude of this information. Figure 1 shows a schematic arrangement for such a system.

In conception, this traditional system is untidy due to asymmetry when handling positive and negative numbers. A simplified positioning system is described in a recent British Patent specification.[1] This has the advantage that all computing is executed directly from a reflected-binary (Gray) code. However, the absolute value of the error signal is not known but a discrete number of control signals are available as the member approaches the desired position.

The use of an odd radix for measurement provides the possibility of making a computing system which handles positive and negative quantities in exactly the same way as well as giving an absolute indication of the relative position, by placing the zero in the centre of the arithmetic values chosen. In the case of a three-valued system, the values used to represent each of the three states are: -1, 0 and $+1$ (abbreviated to $\bar{1}$, 0 and 1). As an example of number representation in this form, Table 1(a) gives the three-valued equivalent of several decimal numbers. These illustrations show immediately one of the important features of the system in that the overall sign of a number is incorporated within each digit of the three-valued representation. The sign of a three-valued number is changed by simply replacing 1 with $\bar{1}$ and $\bar{1}$ with 1 in each digit position. Some of the properties of signed digital values have been reported by Richards[2] and

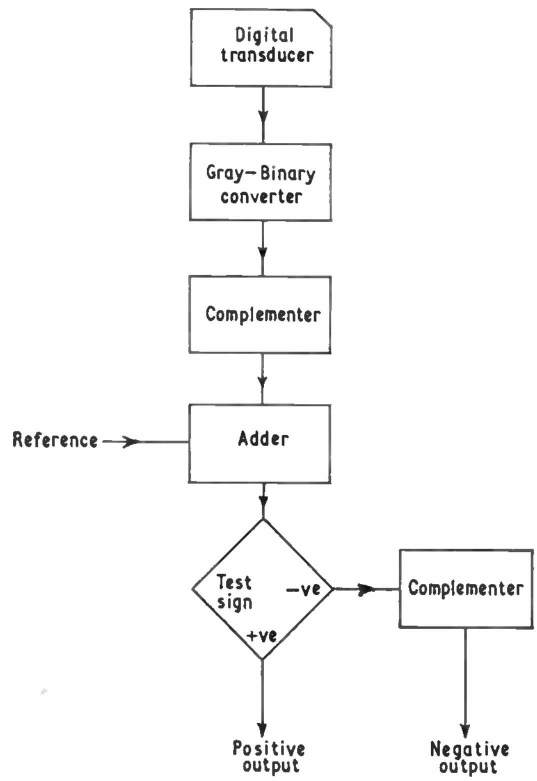


Fig. 1. Schematic arrangement of a typical digital position indicator.

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

Table 1

Examples of three-valued number representation

Decimal value	(a) Three-valued digits				(b) Binary-coded digit pairs							
	D 3^3	C 3^2	B 3^1	A 3^0	D^- -3^3	D^+ $+3^3$	C^- -3^2	C^+ $+3^2$	B^- -3^1	B^+ $+3^1$	A^- -3^0	A^+ $+3^0$
15	1	$\bar{1}$	$\bar{1}$	0	0	1	1	0	1	0	0	0
−15	$\bar{1}$	1	1	0	1	0	0	1	0	1	0	0
19	1	$\bar{1}$	0	1	0	1	1	0	0	0	0	1
−19	$\bar{1}$	1	0	$\bar{1}$	1	0	0	1	0	0	1	0
8	0	1	0	$\bar{1}$	0	0	0	1	0	0	1	0
−8	0	$\bar{1}$	0	1	0	0	1	0	0	0	0	1

Table 2

Truth table defining sum and carry

A	B	Sum	Carry
1	1	$\bar{1}$	1
1	0	1	0
1	$\bar{1}$	0	0
0	1	1	0
0	0	0	0
0	$\bar{1}$	$\bar{1}$	0
$\bar{1}$	1	0	0
$\bar{1}$	0	$\bar{1}$	0
$\bar{1}$	$\bar{1}$	1	$\bar{1}$

Hanson,[3] but it is useful to consider a few examples at this stage. Table 2 gives the truth values for the generation of sum and carry as a result of the addition of two three-valued digits, A and B. With the help of this table, the addition, 15+19 = 34, is performed as follows:

$$\begin{array}{rcccccl} 15 \equiv & 1 & \bar{1} & \bar{1} & 0 & & \\ 19 \equiv & 1 & \bar{1} & 0 & 1 & & \\ \hline & 1 & 1 & \bar{1} & 1 & & \text{sum} \\ & \swarrow & \swarrow & \swarrow & \swarrow & & \\ 0 & \bar{1} & 0 & 0 & & & \text{carry} \end{array}$$

The result is given by:

$$(1\times 3^3)+(1\times 3^2)-(1\times 3^1)+(1\times 3^0)$$
$$= 27+9-3+1 = 34$$

Subtracting 19 from 15 only requires a change in sign in each of the digit values of the three-valued representation of 19 (i.e. to give −19) before *adding* to 15, as follows:

$$\begin{array}{rcccc} 15 \equiv & 1 & \bar{1} & \bar{1} & 0 \\ -19 \equiv & \bar{1} & 1 & 0 & \bar{1} \\ \hline & 0 & 0 & \bar{1} & \bar{1} \end{array}$$

This result has the decimal equivalent of −4.

It is possible to employ binary coding of the three-valued digits in the form given in Table 1(b). Here, the positive and negative component parts of the digit values are separated and appear in columns with a signed weighting. The advantage of using this type of representation is clear, because it is now possible to use binary logic circuits and elements to perform three-valued operations. In the case of this binary-coded system, a change in sign of each digit is realized by transposing the two component parts of a single three-valued digit pair. For digital systems which do not require bulk storage, this form of coding has much to offer in the way of simplification of arithmetic operations.

One of the major features discussed in this paper concerns the evolution of a binary-coded three-valued cyclic (Gray) code together with conversion logic suitable for restoration of true three-valued information. A provisional patent specification has been filed jointly by the National Research Development Corporation and the author under application No. 4558/68, which covers these proposals. The techniques used for the design of three-valued binary-coded logic have been described previously by the author.[4]

2. Three-valued Cyclic Codes

As indicated in the Introduction, no attempt is made to consider practical devices which allow a three-valued code to appear in single digit positions. It is assumed at the outset that all the conventional realizations of such a system would employ analogue techniques. For example, a photo-electric system could be designed using a partially opaque level as well as the clear and completely opaque levels used in a binary system. Discrimination of these three levels would introduce many unwanted problems associated with analogue techniques. Instead, a single three-valued digit will occupy two binary digit positions on the transducing medium of the digitizer. The four possible values of two binary digits introduces a redundant state, and for the purpose of this paper, it will be assumed that the situation where both binary digits have the value 1 never arises. It is evident at this stage that this form of coding is slightly inefficient when compared with the equivalent binary code, but the advantages offered by three-valued codes largely offset this inconvenience.

2.1. *Code Construction*

A cyclic code is best assembled with the aid of a Veitch diagram (or Karnaugh map), which is a 'straight line' version of a Venn diagram. Each small area in a Veitch diagram represents a particular combination of digital values. For example, 16 areas (i.e. 2^4) are necessary to represent the possible combinations of four binary variables. The value of such a diagram for code construction is that adjacent squares or areas in the diagram differ only in the changing of one of the digit values. Phister[5] describes in greater detail how cyclic codes are constructed using Veitch diagrams.

Before proceeding with the construction of a three-valued code, it is necessary to recognize that the redundant combinations mentioned earlier must be accommodated on the diagram. The prohibited states are represented by the shaded area in the Veitch diagram of Table 3 for two three-valued digits. The code sequence, therefore, must be confined to the unshaded portion: each change in the code being displaced from its neighbour by a single square. (Squares at opposite ends of rows or columns are also regarded as adjacent.)

It is convenient to consider just two three-valued digits for code construction as an aid to simplicity at this stage. The elementary observations for such a code are (i) it will have nine combinations, and (ii) it will be symmetrical about zero. It is clearly logical to position the square corresponding to zero at position *t* in the enclosed area which corresponds to the decimal value 0. It is also logical to fix the beginning and end of the code with the digit values corresponding to the decimal numbers -4 and $+4$. These values will correspond with positions *p* and *x* respectively. The code, at this stage of construction is also shown in Table 3. There remain only two possible alternatives for the code sequence and these are given in Table 4. The values defined by Table 4(a) are plotted in Table 5, together with the natural three-valued code, the decimal equivalent, and its position in Table 4(a). An asterisk against a row indicates that the cyclic and natural codes are identical.

A similar exercise with Table 4(b) shows that the resultant code is identical with that of Table 4(a), but with the three-valued digits A and B interchanged. Clearly, Table 4(a) offers the most realistic code. Many other possibilities for codes are apparent if the initial restrictions are relieved, for example, fixing only the position *t* but not positions *p* and *x*. An investigation of these alternatives reveals that other useful codes can be constructed, but they all show some similarity with the codes defined in Table 4.

Table 3

Prohibited area and initial conditions for a binary-coded three-valued cyclic code.

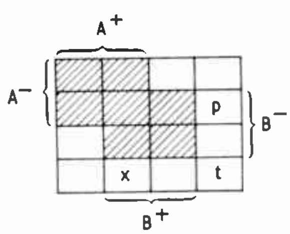


Table 4

Two possible alternative cyclic code configurations

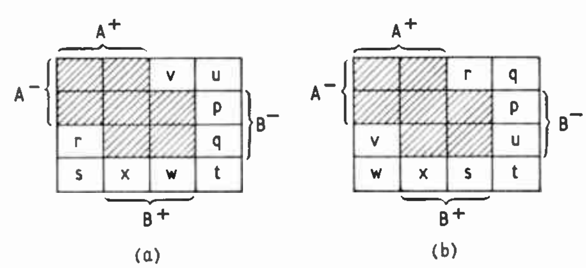


Table 5

The code defined by Table 4(a).

Position in Veitch diagram	Cyclic				Natural				Decimal equivalent	
	B^-	B^+	A^-	A^+	B^-	B^+	A^-	A^+		
p	1	0	1	0	1	0	1	0	-4	*
q	1	0	0	0	1	0	0	0	-3	*
r	1	0	0	1	1	0	0	1	-2	*
s	0	0	0	1	0	0	1	0	-1	
t	0	0	0	0	0	0	0	0	0	*
u	0	0	1	0	0	0	0	1	1	
v	0	1	1	0	0	1	1	0	2	*
w	0	1	0	0	0	1	0	0	3	*
x	0	1	0	1	0	1	0	1	4	*

Returning to Table 5, it is apparent that the only dissimilarity between the natural and Gray versions arises when digit B has the value 0, when the components of digit A are interchanged. However, if a third digit, C, is added, the components of both B and A need to be interchanged when C = 0, with the exception that the components of A do not require interchanging when C = B = 0. Development of this code to a higher number of digits promotes a pattern whereby it is possible to specify the design rules for the necessary conversion logic between cyclic and natural three-valued representation.

2.2. *Code Conversion*

At the outset, it is clearly necessary to provide a logic arrangement to interchange the positions of the component parts of a three-valued digit depending on the presence of an *invert* control signal computed from the value of higher order digits. The truth table defining the logical properties of such an inverter is given in Table 6 for the component parts of a three-valued digit in the cyclic code, namely, X^+_{ig} and X^-_{ig}. I_i is the invert signal delivered to the digit X_i. The redundant states have been omitted from this table.

Table 6

Truth table defining the requirements of conversion logic for a single-digit pair.

X^-_{ig}	X^+_{ig}	I_i	X^-_i	X^+_i	I_{i-1}
0	0	0	0	0	1
0	1	0	0	1	0
1	0	0	1	0	0
0	0	1	0	0	0
0	1	1	1	0	1
1	0	1	0	1	1

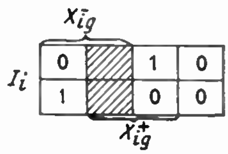
Also shown in Table 6 are the truth values for the invert signal I_{i-1} delivered to lower-order digits. These values have been calculated from the conversion pattern discussed in Section 2.1. This is best summarized by the following expression:

$$I_{i-1} = \begin{cases} 0 \text{ if} \begin{cases} X_i \neq 0 & \text{and} \quad I_i = 0 \\ X_i = 0 & \text{and} \quad I_i = 1 \end{cases} \\ 1 \text{ if} \begin{cases} X_i = 0 & \text{and} \quad I_i = 0 \\ X_i \neq 0 & \text{and} \quad I_i = 1 \end{cases} \end{cases} \qquad \ldots\ldots(1)$$

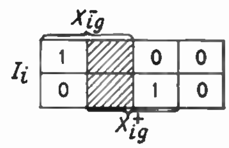
The Boolean expressions resulting from Table 6 are represented by the Veitch diagrams of Table 7 and the associated equations.

Table 7

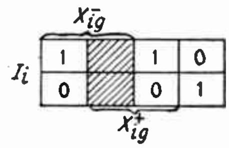
Veitch diagrams and resultant equations from the functions of Table 6



(a) $X^-_i = X^-_{i.}\overline{I_i} + X^+_{ig} I_i$



(b) $X^+_i = X^+_{ig}\overline{I_i} + X \quad I_i$



(c) $I_{i-1} = X^-_{ig} I_i + X^+_{ig} I_i + \overline{X^-_{ig}}\,\overline{X^+_{ig}}\,\overline{I_i}$

If it is assumed that NAND operators are used to construct these functions, the expressions for X_i and I_{i-1} can be written in the following form:

$$X^-_i = \overline{\overline{X^-_{ig}\overline{I_i}} \,.\, \overline{X^+_{ig} I_i}} \qquad \ldots\ldots(2)$$

$$X^+_i = \overline{\overline{X^+_{ig}\overline{I_i}} \,.\, \overline{X^-_{ig} I_i}} \qquad \ldots\ldots(3)$$

$$I_{i-1} = \overline{\overline{X^-_{ig} I_i} \,.\, \overline{X^+_{ig} I_i} \,.\, \overline{\overline{X^-_{ig}}\,\overline{X^+_{ig}}\,\overline{I_i}}} \qquad \ldots\ldots(4)$$

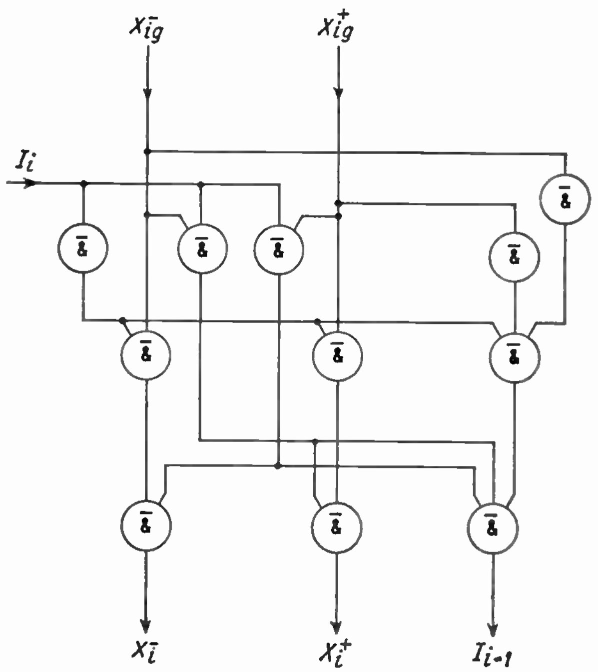


Fig. 2. Code conversion logic for a single binary-coded three-valued digit pair.

The schematic arrangement resulting from these equations is shown in Fig. 2. Many alternatives are possible depending on the availability of logic operators supplied by any given manufacturer. It is interesting to note that the same circuit will generate the cyclic code from the natural form.

3. Design of a Three-valued Full-adder

To compare the magnitude of the digitizer output with some reference, it is necessary to subtract one number from the other. As demonstrated in Table 1, this is basically a process of addition where one number is added to the other with its sign changed. In practice, this change of sign is accomplished by simple commutation of the wires representing the component parts of a single digit.

The essential feature of the logical manipulation is the incorporation of the redundancy or 'don't care' states arising from the binary representation of three-valued logic.

3.1. *Design Requirements*

Investigation has shown that nothing is to be gained by designing a full-adder directly from the truth values, over the construction of a full-adder from two half-adders. The schematic arrangement using half-adders is shown in Fig. 3.

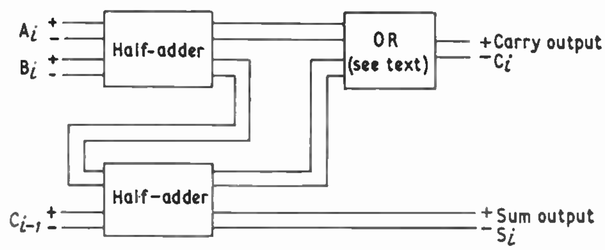


Fig. 3. Schematic arrangement of a full-adder.

The binary component parts of the three-valued digits are used to form a truth table for the generation of the component parts of sum and carry. It is, of course, possible for a carry to be generated in either half-adder, but examination of the truth table defining full addition demonstrates that it is not possible for a carry of similar components to be generated at the same time by both half-adders. However, it is possible for one half-adder to generate a positive carry while the other half-adder generates a negative carry, and vice versa. The combination of the two pairs of carry signals is basically an OR-type operation, with the exception that cancellation is required when opposite polarity signals are present. In fact, the design of this OR element is similar to the design of the half-adder described in the following section with the simplification in the truth table arising from the 'don't care' conditions when two components of the same polarity are contained in a single entry.

Table 8

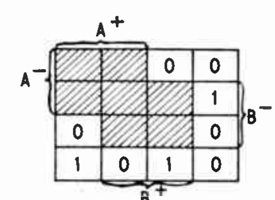
Truth table for the sum and carry functions of a binary-code half-adder

A^-	A^+	B^-	B^+	f^-	f_s^+	f_o^-	f_o^+
0	0	0	0	0	0	0	0
0	0	0	1	0	1	0	0
0	0	1	0	1	0	0	0
0	0	1	1	*	*	*	*
0	1	0	0	0	1	0	0
0	1	0	1	1	0	0	1
0	1	1	0	0	0	0	0
0	1	1	1	*	*	*	*
1	0	0	0	1	0	0	0
1	0	0	1	0	0	0	0
1	0	1	0	0	1	1	0
1	0	1	1	*	*	*	*
1	1	0	0	*	*	*	*
1	1	0	1	*	*	*	*
1	1	1	0	*	*	*	*
1	1	1	1	*	*	*	*

The positive and negative components of the sum and carry functions are f_s^+, f_s^-, f_o^+ and f_o^- respectively. An asterisk indicates a redundant combination.

Table 9

Veitch diagram for the positive component of the sum function of Table 8



3.2. *Detail Design of Half-adder*

The techniques used for the design of a binary-coded adder are described in detail in a previous publication by the author.[4]

The truth table defining this operation is given in Table 8, with the positive component of the sum output shown in the Veitch diagram of Table 9. Making use of the 'don't care' states, the simplest Boolean expression for this function is

$$f_s^+ = A^- B^- + A^+ \overline{B^+}\,\overline{B^-} + \overline{A^-}\,\overline{A^+} B^+ \quad ......(5)$$

Similarly, the negative component of the sum function can be shown to be

$$f_s^- = A^+ B^+ + \overline{B^+}\,\overline{B^-} A^- + \overline{A^+}\,\overline{A^-} B^- \quad ......(6)$$

The positive and negative components of the carry function are simply given by

$$f_c^+ = A^+ B^+ \qquad ......(7)$$

$$f_c^- = A^- B^- \qquad ......(8)$$

If NAND logic is used, eqns. (5) and (6) can be written as:

$$f_s^+ = \overline{\overline{(A^- B^-)}\,\overline{(A^+ \overline{B^+}\,\overline{B^-})}\,\overline{(A^+ \overline{A^-} B^+)}} \qquad ......(9)$$

$$f_s^- = \overline{\overline{(A^+ B^+)}\,\overline{(A^- \overline{B^+}\,\overline{B^-})}\,\overline{(A^+ \overline{A^-} B^-)}} \qquad ......(10)$$

A possible realization for the half-adder is shown in Fig. 4 using the expressions defined by the above four equations.

This realization of the sum and carry functions differs from the proposals in Ref. 4 due to the fact that it is virtually certain that integrated circuits will be used in the construction in association with two-wire transmission of three-valued information. Sufficient information is contained in this Section for the design of a full-adder. The two-wire representation may be a valuable asset for decoding the output of the adder into positive and negative analogue signals for operating actuators.

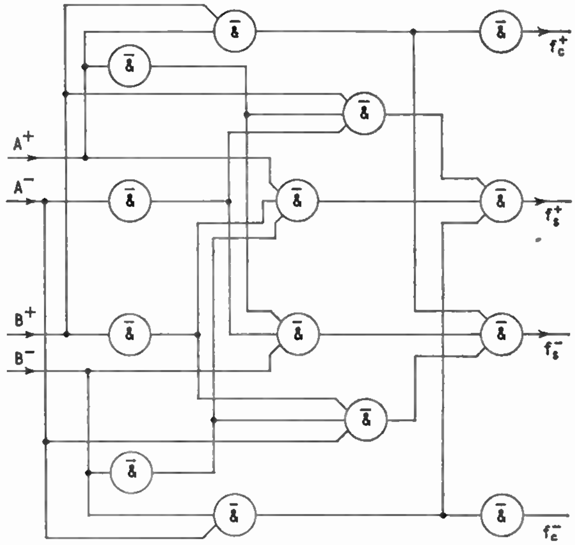


Fig. 4. Logic diagram of half-adder using NAND connectives.

4. Conclusions

In a comparison with a conventional binary system, the few disadvantages of a three-valued system are largely offset by the simplified computing system required. In the case of the digitizer, eight binary tracks are required to give a resolution of one part in 256. An equivalent resolution using a three-valued binary coded transducer requires 10 binary tracks to give a resolution of one part in 243. This is, perhaps, the major disadvantage. However, it must be remembered that the incremental size of the least significant digits will be the same in both systems, and this is the limiting factor in the design of digital transducers. The remainder of the system offers many attractive advantages. The number of gates required for a single three-valued digit is very similar to the number required for two binary digits when the code converter and full-adder are considered alone. However, in a binary system, complementing and decision circuits are required, in addition to the code converter and full-adder, which are not necessary when three-valued coding is used. The need to compute to one extra digit to determine sign, which is necessary when using complemented binary numbers, does not arise in the case of three-valued systems.

A further advantage is apparent when digital-to-analogue (d.-a.) conversion is considered The positive and negative component parts of each three-valued digit are available separately and they can be used as such to switch signed, weighted currents in the analogue summing circuits of a d.-a. converter. Taking the numerical example above, a total of 16 switches would be necessary to handle both polarities in an eight-digit binary system, whereas, only 10 switches would be necessary in a binary-coded three-valued system, for comparable accuracy.

If the manufacturers of integrated circuits made available single packages of, say, three-valued half-adders, then the advantages offered by a three-valued system would be even more obvious, providing a valuable saving in the cost and complexity of positioning systems of this type.

5. Acknowledgments

The author would like to thank the University of Nottingham for providing the facilities to carry out this work and the National Research Development Corporation for sponsoring the related Patent Application.

6. References

1. Unimation Inc. British Patent No. 1,050,584.
2. Richards, R. K., 'Arithmetic Operations in Digital Computers' (Van Nostrand, New York, 1955).
3. Hanson, N. H., 'Ternary threshold logic', *Trans. Inst. Elect. Electronics Engrs on Electronic Computers*, **EC-12**, pp. 191–7, June 1963.
4. Pugh, A., 'Application of binary devices and Boolean algebra to the realisation of 3-valued logic circuits', *Proc. Instn Elect. Engrs*, **114**, pp. 335–8, March 1967.
5. Phister, M., Jr., 'Logical Design of Digital Computers', pp. 232–4 (Wiley, New York, 1958).

Manuscript first received by the Institution on 27th May 1968 and in final form on 23rd July 1968. (Paper No. 1214/IC3.)

U.D.C. 621.397.132.3

Colour Television Studio Equipment and Problems

By

F. G. PARKER,

B.Sc., C.Eng., M.I.E.E., M.I.E.R.E.†

Presented at a Meeting of the Institution's Communications Group held in London on 24th January 1968.

Summary: The paper outlines the nature of a studio complex and the problems which arise from the introduction of colour television beginning with those associated with the colour separation signals. Special reference is made to the problems arising from the use of four-tube colour cameras particularly from the point of view of monitoring their outputs as well as the methods of coding to ensure accurate colour reproduction on commercial receivers. In a studio using the PAL system of colour television certain extra pulses are required for use with the coders and the generation and distribution of these pulses is discussed.

The problems arising in the coding process are dealt with and in particular special consideration is given to the problems due to the colour axis switching feature of the PAL system. The effect of the axis switching on the increased precision needed in the quadrature between the two axes is also considered as well as its effect on the crosstalk problems. The design of a practical operational coder is outlined with particular reference to the problems of carrier balance and the effect of the axis switching on this parameter.

The effect of colour on the mixing equipment is considered and particular reference is made to the problem of burst-stabilization during the operation of the fader controls. The subject of special effects is also discussed including the production of synthetic colour captions from black and white originals. The synchronization of the various sources at the mixer is considered in outline in so far as it affects the mixer.

The design of a suitable decoder for use with colour picture monitors in a colour studio is outlined and also the general arrangements for colour picture monitoring. The requirements of a signal distribution system within a studio complex are discussed with reference to such parameters as differential gain and chrominance luminance delay.

Finally, the paper considers the method of line-up necessary to ensure that coding, decoding and monitoring equipment is working correctly. In particular, a method is outlined whereby coding equipment can be completely set up using an oscilloscope to monitor the output of a coder.

1. Introduction

Although the introduction of colour to a television service brings new equipment, new techniques and new problems, the fundamental function of a colour television studio complex remains the same as that of its monochrome counterpart, namely to produce a composite programme having entertainment value. In order to achieve this object the light from the studio scene must be turned into an electrical signal which is then subjected to various operations such as fading, mixing and routing before being applied to the transmitter for distribution to the audience. Associated with these operations are those of signal monitoring and equipment line-up to ensure an adequate standard of output. Because the camera considerations are very adequately dealt with elsewhere,[1] in this paper the signal is considered as beginning at the output of the picture source, such as the camera, and the various processes between the camera output and the output of a studio complex are considered.

To produce a television programme of half an hour's duration it is necessary to present, in a time sequential fashion, the outputs from a number of cameras so disposed within the studio that the mood of the drama, for instance, can be captured by the method and timing of the interchanges between the cameras. Added depth can be given to the programme by the use of inserts which may, for example, have been filmed on location and replayed on a telecine machine for combining into the programme. Sometimes material is pre-recorded on video tape and arrangements are necessary to insert this type of

† Designs Department, British Broadcasting Corporation, London, W.1.

material into the programme. In some programmes such as those dealing with current affairs, the producer often wishes to combine the outputs of his local cameras with the outputs of some cameras situated some hundreds of miles distant, and special equipment is necessary to allow this facility.

The change between picture sources can be made in three distinct ways; namely by a cut, a cross-fade, or a wipe. The method chosen is dependent on the situation, for instance, the news programmes mainly use cuts, drama uses both cuts and fades, the cuts for fast action and the fades to indicate a change of venue or time. Light entertainment programmes use all three modes.

To produce a network programme such as BBC-2, for example, the outputs of studios, telecines, and video tape recorders are taken and likewise combined time sequentially. The central control mixers perform this function but working on an extended time scale compared with the studio mixer, the method of working being the same.

The application of colour to a programme must not limit the facilities available to a producer nor must the conditions of use of the facilities be so restrictive that the operational routines become unworkable; therefore the facilities of a colour studio must be the same, and be capable of being operated in the same way as those of a black-and-white studio. Colour, even with television, is really only a part of the scenery as it is with film.

This had led to the adoption of the philosophy that handling colour in a studio complex is just an extension of black and white television, but in some cases requiring more equipment.

2. Coded Versus RGB Working

Original picture sources such as cameras and telecines, both of which can be either of the three-tube or four-tube type, give out a number of signals either all wideband in the case of a three-tube device or one wideband signal together with three narrower bandwidth signals in the case of a four-tube device. The signals appearing at the camera outputs in this form are known as RGB signals in the case of the three tube camera, or RGBY for the four-tube camera. At this stage the broadcasting authority can make the decision whether to perform all the mixing operations and distribution with the signals in these forms or code them and use coded mixing.

To undertake RGB working, for example, the black and white equipment throughout a studio complex must be triplicated, but only one coder per studio complex would be required with such a mode of working. Certain difficulties arise in the case of sources remote from the studio as they would require decoding to RGB before they could be included in the final programme. The case of RGBY working is not considered since as is shown in another section of the paper these signals can, by simple matrixing, be reduced to suitable RGB signals.

If coded mixing is decided upon then each source is converted from RGB to coded at a very early stage. This operation results in the amount of equipment necessary for mixing and distribution being the same as for the black-and-white studio. In the B.B.C. coded working has been adopted.

The decision to use coded mixing was not entirely based on the amount of equipment being less but on the fact that signals in their RGB form are somewhat delicate and need handling very carefully. It is true, of course, that they are not affected by such an exotic distortion phenomenon as differential phase distortion but they can be irreparably damaged by the more mundane distortions. For example, the camera is arranged to give out equal amplitude signals on white and from this all other colours follow. If in the process of handling the RGB signals the equality of gain through the system for all three channels is not maintained then colour errors result. Variations of gain as routes change, due to change in the termination resistors for each channel, is significant. Likewise the three channels must have transfer characteristics which are as near identical as possible and overload points must coincide where this is appropriate. Route lengths must be equal in time for all the signals otherwise the care taken in camera registration is wasted, since such a timing error will give rise to coloured fringes all over the picture on vertical detail. Once the signals are coded these difficulties disappear and are replaced by differential phase and differential gain distortion which need particular care and attention but are readily dealt with.

There are, however, certain areas where the colour signal must be handled in its RGB form and this is between the output of the camera control unit and the coder input. In this path are usually some distribution amplifiers to allow a feed of the camera output, as RGB, to be applied to a colour monitor for picture matching, as well as to the colour monitor with the camera control unit and the coder. The variation in relative gains for R, G and B in this path could be due to changes in amplifier gain, terminating resistances, or source resistances. The first variation, namely that of gain, is taken care of by having amplifiers with a large amount of negative feedback and the remaining two by using close tolerance resistances for the output impedances and the terminating impedances of the amplifier. The resistance tolerances chosen are $\pm 0{\cdot}2\%$ giving a level variation of $0{\cdot}2\%$ worst case due to these causes. When using distribution amplifiers having several outputs it is essential that if the three channels are adjusted to be equal at No. 1 output, for example, then this equality must be maintained for all other outputs. The method of using close tolerance resistances for both source and terminating resistors provides the easiest and, in the long run, the cheapest solution to the problem.

3. Camera Coding and Monitoring

3.1. *General Considerations*

Having decided to code the signals at the earliest opportunity the situation in the case of the three-tube camera is straightforward, the signals are fed to a coder; but the four-tube camera still leaves the broadcaster with two questions; firstly, can its output be adequately monitored without the use of a coding and decoding system and secondly, what is the best coding method to use with a four-tube camera. To understand the problem a little further a look at the composition of the signals of the colour system used is advantageous. Basically, a colour system is a data transmission system conveying the three colour separation signals from the camera to the tricolour tube grids to reproduce light to simulate the studio scene. To achieve this every colour system uses a luminance signal and chrominance signal with the object of producing from these two signals at the receiver the three signals $R^{1/\gamma}$, $G^{1/\gamma}$, $B^{1/\gamma}$ to apply between the cathodes and grids of the display device.

3.2. *Three-tube Camera Case*

The three-tube camera outputs which are $R^{1/\gamma}$, $G^{1/\gamma}$, and $B^{1/\gamma}$ are matrixed to produce a luminance signal $Y' = 0{\cdot}3R^{1/\gamma}+0{\cdot}59G^{1/\gamma}+0{\cdot}11B^{1/\gamma}$, and two chrominance signals $R^{1/\gamma}-Y'$, and $B^{1/\gamma}-Y'$. At the receiver the addition of Y' to the two chrominance signals gives $R^{1/\gamma}$ and $B^{1/\gamma}$. The signal $G^{1/\gamma}$ is recovered by using the fact that Y' is a function of $R^{1/\gamma}$, $G^{1/\gamma}$, $B^{1/\gamma}$. The three-tube camera therefore gives correct colour reproduction but the black-and-white viewer does not get a true monochrome picture since Y' is not the same as $Y^{1/\gamma}$ which is

$$(0{\cdot}3R+0{\cdot}59G+0{\cdot}11B)^{1/\gamma}.$$

The difference between these two quantities is greatest for low luminance colours.

3.3. *Four-tube Camera Case*

The outputs of the four-tube camera are $R^{1/\gamma}$, $G^{1/\gamma}$, $B^{1/\gamma}$ and $Y^{1/\gamma}$. If the luminance signal $Y^{1/\gamma}$ and the chrominance signals $R^{1/\gamma}-Y^{1/\gamma}$, $B^{1/\gamma}-Y^{1/\gamma}$ were transmitted then it would be possible to recover $R^{1/\gamma}$ and $B^{1/\gamma}$ but the recovery of $G^{1/\gamma}$ would be an involved process. Unless $G^{1/\gamma}$ can be recovered correctly the colour picture suffers so this method of transmission is ruled out. Another method of coding, attributed to Livingston, transmits the luminance signal $Y^{1/\gamma}$ with two chrominance signals $R^{1/\gamma}-Y'$, $B^{1/\gamma}-Y'$ where Y' is made by matrixing the outputs of the three colouring tubes. This method again does not produce the correct result but one which is more acceptable. The receiver can recover $R^{1/\gamma}-Y'$, $B^{1/\gamma}-Y'$ and $G^{1/\gamma}-Y'$ but can only add to these $Y^{1/\gamma}$ which never results in $R^{1/\gamma}$, $G^{1/\gamma}$, $B^{1/\gamma}$ being applied to the tube. This results in errors in the colour picture particularly when $Y^{1/\gamma}$ departs appreciably from Y'.

The advent of plumbicon tubes, with a linear output characteristic, has provided the opportunity to look into other ways of dealing with the outputs of a four-tube camera. Since the object of colour television is to produce good colour pictures it is obvious that within the confines of the system and the complexity of the receiver this can only be done if the luminance signal transmitted is actually Y' over the bandwidth of the colour information. This condition can be achieved by the use of a classic matrix which converts $Y^{1/\gamma}$ into Y' over the bandwidth of the chrominance signals. To achieve this a correction signal known as the ΔL correction signal,[2] which is the difference between $Y^{1/\gamma}$ and Y', is produced within the matrix and subtracted from the $Y^{1/\gamma}$ signal. The resulting signal $Y^{1/\gamma}-\Delta L$ is equal to Y'.

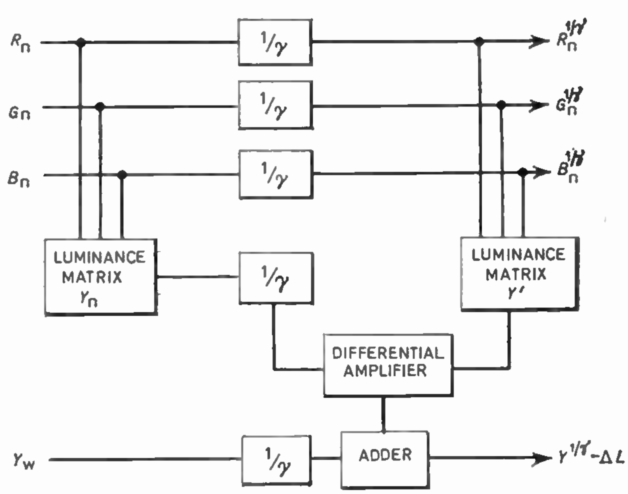


Fig. 1. Classic matrix.

The layout of a classic matrix is shown in Fig. 1. In such a matrix the key items are the gamma correctors which must track over a large range and maintain their performance over an adequate temperature range. The output signals from the matrix, which are the $Y^{1/\gamma}-\Delta L$ signal together with the narrow band $R^{1/\gamma}$, $G^{1/\gamma}$ and $B^{1/\gamma}$ signals, can be applied to a coder and transmitted so that a true colour picture results.

3.4. *Four-tube Camera Monitoring*

Just as there was a problem with coding a four-tube camera the problem of monitoring it can be solved in two ways; the pictures can be either monitored via a coding and decoding system or by producing equivalent RGB output signals. The latter method has been adopted and a matrix has been built to convert the three narrow-band outputs and the one wideband output into three wideband outputs; that is, convert the four-tube camera to a three-tube camera as regards the electrical outputs. Such a matrix consists of a number of phase splitters, one for each colouring signal, together with resistive adding networks. The incoming wideband luminance signal can come either direct from the four-tube camera or via a classic matrix. If the narrow band chrominance signals are written as R_n, G_n and B_n and the wideband luminance signal as Y_w then the equations relating the outputs of the matrix to inputs are as follows:

$$R_o = R_n - Y_n + Y_w$$

$$G_o = G_n - Y_n + Y_w$$

$$B_o = B_n - Y_n + Y_w$$

when Y_n is $0{\cdot}3R_n + 0{\cdot}59G_n + 0{\cdot}11B_n$.

At frequencies above the bandwidth of the colouring channels $R_n = G_n = B_n = 0$ hence $Y_n = 0$ and therefore each output is Y_w. Such a matrix basically produces a mixed-highs type of signal. Over the bandwidth of the colouring channels the three outputs are R_n, G_n and B_n respectively if $Y_n = Y_w$ which will be the case if Y_w has been modified to Y' by adding ΔL in a classic matrix.

Although this linear matrix in principle simulates the action of a coding and decoding system it does not reproduce their action exactly since it does not produce intermodulation products between high-frequency noise and the sub-carrier. Therefore its use, particularly in comparison work, should be carefully watched. In addition, although signals can be added and subtracted, noise only adds, so there is a deterioration of the signal/noise ratio.

When using a four-tube camera the output can be coded in the Livingston form or in the normal form by the use of a classic matrix. By use of the matrix described above the four-tube camera outputs can be converted to wideband RGB signals and then treated, both for coding and monitoring, as a three-tube camera.

4. Ancillary Signals

Before the colour signals coming from a camera, or telecine, can be mixed they must be applied to a coder which in turn requires certain ancilliary signals not found in a black-and-white studio. These signals are sub-carrier, burst gating pulse, and V-axis switching signal. Since the sub-carrier in the PAL system[3,4] must be locked rigidly to the line frequency, experience has shown that the generation of the sub-carrier is best associated with the waveform generator located in a central area, and distributed to various coder locations. The other two signals are also derived from the waveform generator. These signals require a further distribution system but due to the complexity of the division ratio in the case of the sub-carrier and the blanking in the case of the burst gating pulse the decision to generate them centrally is, in the author's opinion, the correct one.

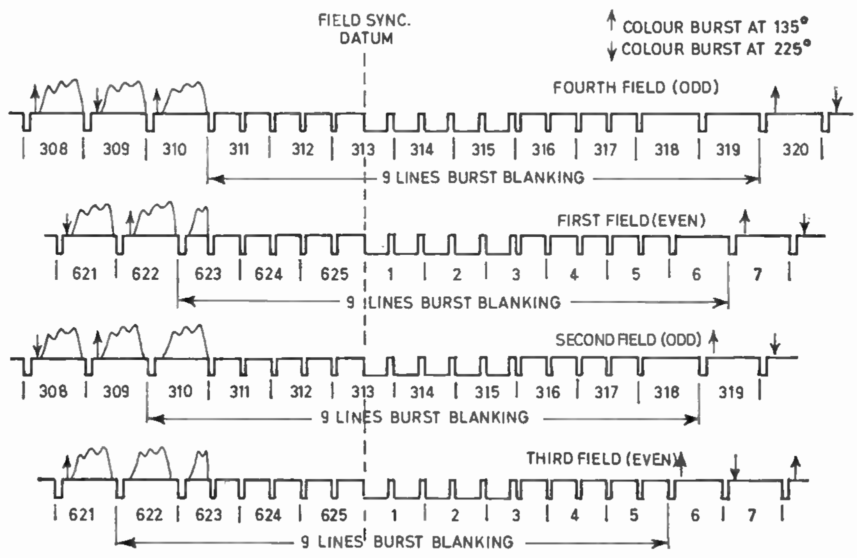


Fig. 2. Burst blanking.

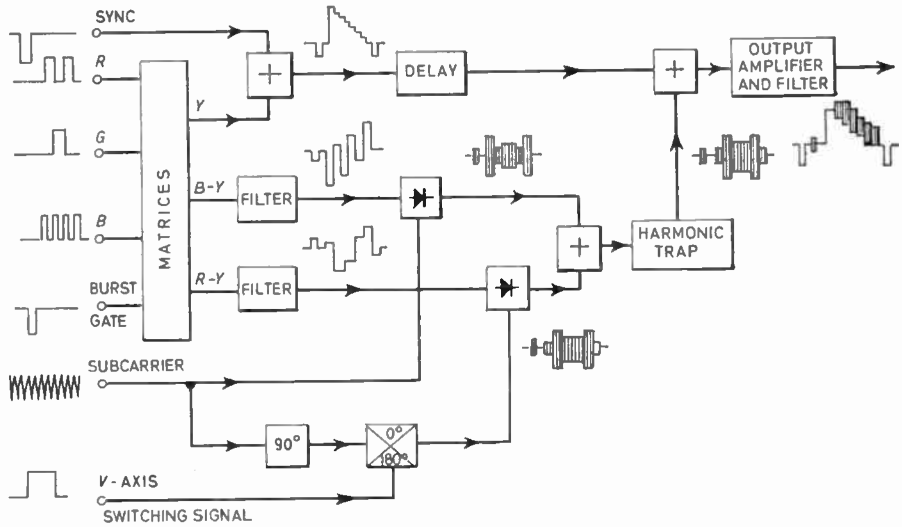


Fig. 3. Block diagram of a PAL coder.

The sub-carrier for the PAL system is derived by taking a frequency which is ($\frac{1}{2}\times 567$) times line frequency and a frequency of $\frac{1}{4}$ line frequency plus 25 Hz is modulated on to it giving the PAL sub-carrier frequency of ($\frac{1}{4}\times 1135$) times line frequency plus 25 Hz, which is 4·433 618 75 MHz. Sub-carrier is fed round the studio complex at a level of 1 V peak-to-peak into 75 Ω and is handled by video distribution amplifiers: the choice of this level was made to reduce crosstalk.

The *V*-axis switching signal which is an output signal from the waveform generator, is a half-line frequency square-wave of 1 V peak-to-peak amplitude, arranged to have its vertical edges coincident with the synchronizing pulse timing edges. The convention which has been adopted in B.B.C. studios is that the positive half-cycle corresponds to the line on which the + *V*-axis is transmitted.

The burst gating pulse is arranged so that its timing with respect to the synchronizing pulses is that of the final burst at the coder output. Prior to the introduction of the blanking of this particular signal, known as 'Bruch blanking', it was possible to generate the burst gating pulse quite easily but now this type of blanking of the burst requires so much extra circuitry that central generation is imperative. In Fig. 2 is shown four fields of the output burst from a coder; the direction of the arrow indicates a burst of 135° in the upward direction and 225° in the downward direction. The significant feature of the burst blanking is that the bursts before and after blanking have the same polarity on every field; also the blanking is always for nine lines. This arrangement, however, does give rise to some lines not having a burst but having picture information.

5. PAL Coder

5.1. *General Considerations*

Having obtained the colour analysis signals and the ancillary signals the colour signals can be coded. The design of a coder for the PAL system was begun by the modification of an N.T.S.C. coder but it was quickly apparent that greater care was necessary with nearly every parameter. With N.T.S.C. very often a distortion was present at the coder but was not obvious; however, with PAL, the switching of the *V*-axis causes certain errors to be blatantly obvious.

The block diagram of a coder is given in Fig. 3 which shows the main functions to be carried out in order to convert the colour analysis signals into a coded signal conforming to the equations for the PAL system which are given below:

$$\text{Line } n: \quad E_M = E_{Y'} + E_U \cos \omega t + E_V \sin \omega t$$

$$\text{Line } n+1: \quad E_M = E_{Y'} + E_U \cos \omega t - E_V \sin \omega t$$

where

$$E_{Y'} = 0{\cdot}299E_R^{1/\gamma} + 0{\cdot}587E_G^{1/\gamma} + 0{\cdot}114E_B^{1/\gamma}$$

$$E_U = 0{\cdot}493(E^{1/\gamma} - E_{Y'})$$

$$E_V = 0{\cdot}877(E_R^{1/\gamma} - E_{Y'})$$

Basically, a coder can be divided into two parts, luminance and chrominance. The luminance part consists of amplifiers and a delay network which is used to delay the luminance signal by an amount equal to the time taken by the chrominance signal in passing through the band-limiting filters and modulators. In the chrominance part of the coder are two chains, one handling the *U*-component and the other the *V*-component. Each chrominance component is handled in the same way. The video signals, *U* and *V* derived from the incoming R, G, B signals are

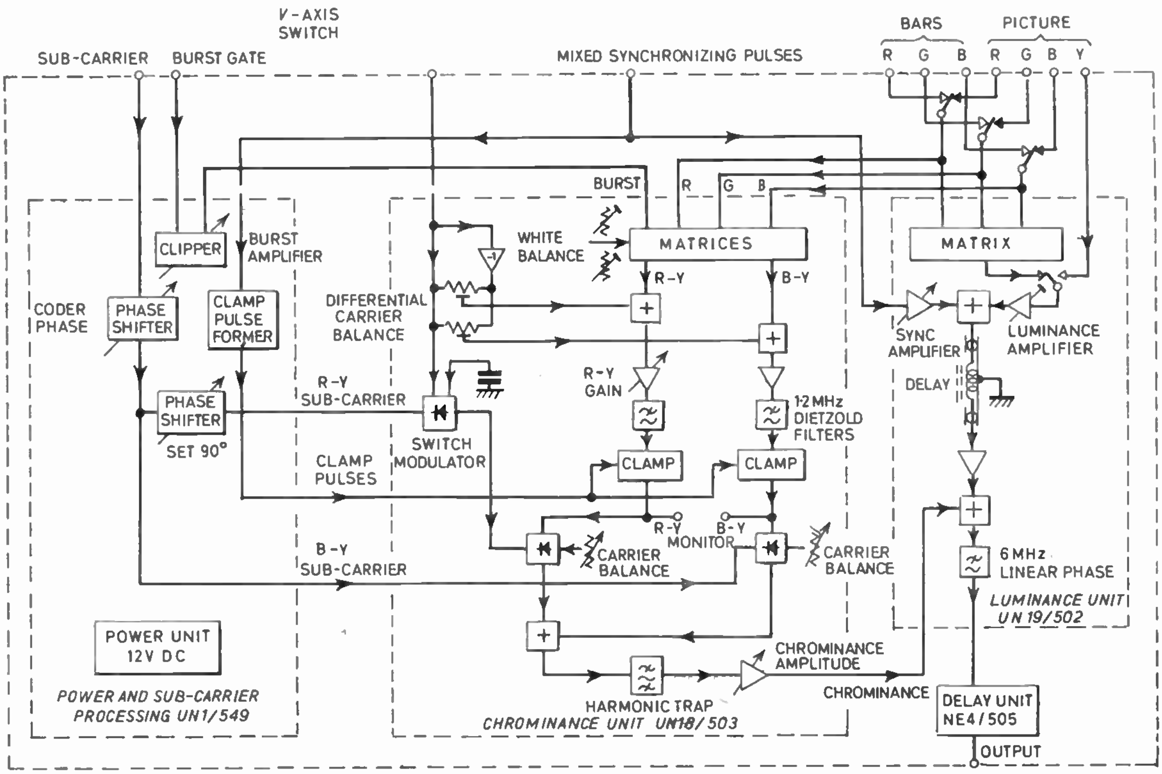


Fig. 4. PAL colour coder (B.B.C. type GE1M/526).

filtered by Dietzold filters to a bandwidth of 1·2 MHz before being applied to the modulators. Since the PAL system uses suppressed carrier amplitude modulation to transmit the chrominance signal the modulators used in the coder are of the ring type which have a good intrinsic balance. The switching of the *V*-axis is accomplished by switching the sub-carrier fed to the *V* modulator rather than the polarity of the *V* video signal due to the difficulty in maintaining the black level sufficiently accurate from line to line. A more detailed block schematic of a coder now in use is shown in Fig. 4.

5.2. *Specific Design Problems*

The problems encountered in the design of such a coder lie in specific points. In the luminance channel the matrix must be accurate and stable, and this is achieved by the use of metal-film resistors of very close tolerance. The delay must be wideband and have a very good pulse response. Failure to achieve this will result in pictures often with a displaced high frequency echo.

The main design problems lie in the design and layout of the chrominance circuitry and arise from, and are highlighted by, the *V*-axis switching. According to the equation of the system the chrominance signal must be zero for areas of black-and-white, this requires that no sub-carrier is transmitted. To achieve this the matrixing of the R, G and B must be such that when they are all equal or zero then there is no input to the modulators. Having achieved this the balance of the modulators themselves must be very good and a figure of 60 dB has been achieved giving an output on black-and-white at the coder output of at least 50 dB. The difference between these two figures is due to inherent crosstalk in the circuitry and it is this crosstalk which causes special problems with PAL. With a system such as N.T.S.C. the crosstalk still existed but was not obvious because the carrier balance of modulators was offset, often unconsciously, to reduce the residual carrier leak to zero on black and white. In the PAL system this technique cannot be easily applied because the action of reversing the phase of one axis from line to line gives rise to a condition where it is not possible to cancel the crosstalk since a different offset would be required for two adjacent lines. A novel way of overcoming this has been devised in which some of the *V*-axis switching square wave is fed to each modulator together with the video. This arrangement gives a different balancing bias from line to line and maintains the carrier balance within the limits at all times.

With the PAL system every colour, except those exactly along the U-axis, has two angles relative to the U-axis, one for each position of the V-axis switch and these angles are symmetrical about the U-axis.

If this condition is not obtained desaturation would appear on Delay PAL receivers and Hanover bars on Simple PAL receivers as well as a colour error. This angular symmetry is dependent on two items of the PAL coder, the accuracy of the 90° between the two axes and the accuracy of the 180° reversal of the V-axis. If the angle between the axes is $(90-\beta)^\circ$ and the V-axis reversal is only $(180-\alpha)^\circ$ then the angular asymmetry between line n and line $n+1$ is given by

$$\phi = \tan^{-1} \frac{UV\cos\beta - UV\cos(\alpha-\beta) + V^2\sin(\alpha-2\beta)}{U^2 + UV\sin(\alpha-\beta) + UV\sin\beta + V^2\cos(\alpha-2\beta)}$$

Substitution shows that ϕ can be as much as 3° for an angular error of 1° in both the quadrature and axis switching. To reduce this error to a minimum, great care is taken to ensure that the V-axis switching is exactly 180° by using a balanced precision modulator as the switch element. Also the 90° between the two axes is made stable and capable of being adjusted to be correct.

Observation of the waveform containing such an error gives the characteristic amplitude twitter of the sub-carrier. Severe amount of this distortion can lead to the Hanover bar pattern on the black-and-white picture.

The equations for U and V show them to be related to R, G, and B as follows:

$$U = 0{\cdot}493(-0{\cdot}3R^{1/\gamma} - 0{\cdot}59G^{1/\gamma} + 0{\cdot}89B^{1/\gamma});$$

$$V = 0{\cdot}877(0{\cdot}7R^{1/\gamma} - 0{\cdot}59G^{1/\gamma} - 0{\cdot}11B^{1/\gamma});$$

$U = V = 0$ when $R = G = B$, which is the case when the scene is black or white. To produce U and V requires the production of negative values of R, G and B and these are obtained by inverter stages. Actually $-U$ and $-V$ are formed in the matrix to reduce circuitry since only one coefficient in each equation is negative. The design of this inverter stage in the U and V channels requires some care, particularly as regards the time delay through it. In early coders the delay between the direct and inverted signals was some 14 ns at the matrix point which gave rise to colour fringing on certain types of pictures, the grapefruit on the 'Bowl of Fruit' test slide being a specific example. This defect was removed by designing the inverter stage to have a time delay of 5 ns relative to the direct path to the matrix.

This coder is used to handle camera and telecine outputs, suitable arrangements being made for the case of four-tube cameras by providing a switch position to accept a separate luminance signal. The sub-carrier fed to the coder is via a phase shifter to assist the colour synchronization of any one coder with any other. In addition the sub-carrier feed to a block of coders is also via a phase shifter so that the phase of a number of coders can be simultaneously moved to bring them into phase at some mixing point with some other signal. To enable the coder to be aligned a switch on the input allows the RGB inputs to be connected to the output of a colour bar generator.

6. Vision Mixing Equipment

The point has now been reached where the outputs from the cameras and telecines have become coded signals and they can now be considered for mixing.

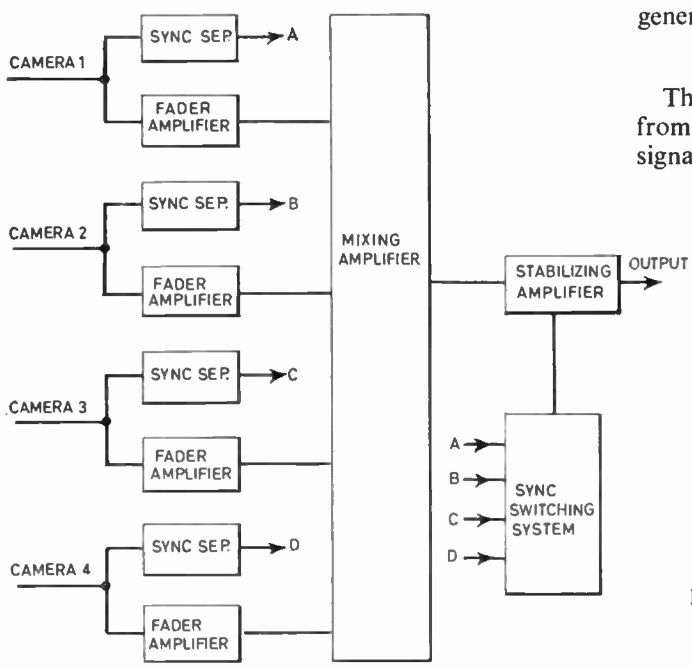


Fig. 5. Schematic of a 'knob-a-channel' type of mixer.

The other coded signal in a studio centre is the one originating from a video tape recording (v.t.r.) machine.

In B.B.C. studios the 'knob-a-channel' mixer predominates and for colour working it brings more problems than an A/B type. The schematic of a mixer is shown in Fig. 5.

A producer can make his programme by either cutting or cross-fading between sources but the cross-fading is limited by the fact that the two sources involved must be synchronous. For black-and-white this requires that the synchronizing pulses of the sources concerned must be coincident within 50 ns, but for colour the two sources between which the the cross fading is to take place must be colour synchronous as well, and this requires that the chrominance sub-carriers be phased to within 1°. The method of achieving synchronism is to use NATLOCK which has been described elsewhere.[5] Associated with the mixer in present black-and-white practice are detectors, one for each mixer input, which check on the synchronizing pulses of the signal applied to each mixer input and decide whether the sources being cross-faded or used for special effects are synchronous. If the sources being used are not synchronous the outputs from the detector, which are applied to an interlock on the fader control, prevent cross-fading occurring when the faders are operated and cuts over to the source. With colour the sources must additionally be checked for colour synchronism and on all input channels of the mixer are placed detectors which check the burst at each input relative to the local burst. If the two bursts are identical in phase and frequency then the two sources can be cross-faded, if not, then cutting only is permitted. The limit set on the phase between two sources is about 5°, beyond this the source is declared non-synchronous.

When operating a mixer some check must be kept on the state of colour synchronism of the various sources applied to the mixer inputs. By synchronism is also implied the phasing of the sources. For operational purposes the best way of checking the colour phasing of sources is to use a polar display such as a vectorscope. This instrument gives a polar display of the state of the phasing and can be interpreted very quickly. The mode of operating the vectorscope is to use it with local studio sub-carrier and then feed the various sources to its inputs via a switching system. When handling colour pictures, as opposed to colour bars, the only recognizable features of each source are the bursts. Generally, there are facilities to allow two inputs, designated A and B, to be displayed simultaneously and using this method phasing can be checked. When operating in this mode the pattern becomes rather confused so the B.B.C. has devised a special polar display which removes the axis switching before display and therefore gives a display similar to that obtained from an N.T.S.C. signal.

Concerning the differential phase distortion and accuracy of source phasing as well, one may ask why bother about this when PAL is being used. There are two answers to this question: the first is that the B.B.C. also operates on 525-line N.T.S.C. for the American market and consequently, in general, studio equipment must be multi-standard; the second is that although the radio industry have said that all receivers will be Delay PAL it seems inevitable that there will be Simple PAL receivers in a few years time if even only a part of the market. The other problems in a mixer arising from the introduction of colour, apart from the differential phase and gain of the mixer, are its overall timing, mainly associated with the stabilizing amplifier.

6.1. *Stabilizing Amplifier*

The stabilizing amplifier is used with a black-and-white mixer to ensure that at all times the mixer gives out synchronizing pulses at the correct amplitude independent of the fader setting; the schematic of the stabilizing amplifier is shown in Fig. 6. Since the signal is handled composite in B.B.C. studios, the policy is to remove the synchronizing pulses in the mixer and replace them with a new set correctly shaped and timed. Their removal is achieved by gating out the synchronizing pulses rather than clipping them off, which would be out of the question with colour signals. The gating pulses for this operation are derived from the synchronizing pulses derived from the incoming signal to the mixer and applied to the diode bridges which, together with the series resistors, form attenuators during the period of the pulses. The width and position of the gating pulses must be such that the burst is not affected. The attenuation of the synchronizing pulses is about 40 dB at this point in the stabilizing amplifier. A new set of synchronizing pulses is inserted near the output. When colour is considered, the synchronizing signal for the chrominance information is the burst and the action of fading will cause it to vary in amplitude just as is the case for the synchronizing pulses. If this is allowed to happen then there could be trouble with the colour synchronization of the colour receiver; for instance, it could lose colour lock, or if it had chrominance a.g.c. the saturation could become excessive, due to a reduced burst amplitude, therefore amplitude stabilization is required for the burst. Apart from amplitude variations of the burst it is also possible that it may be in the wrong position relative to the synchronizing pulses and if this is so then the defect ought to be corrected.

The method of burst stabilization adopted is to compare the burst coming out of the fading amplifier

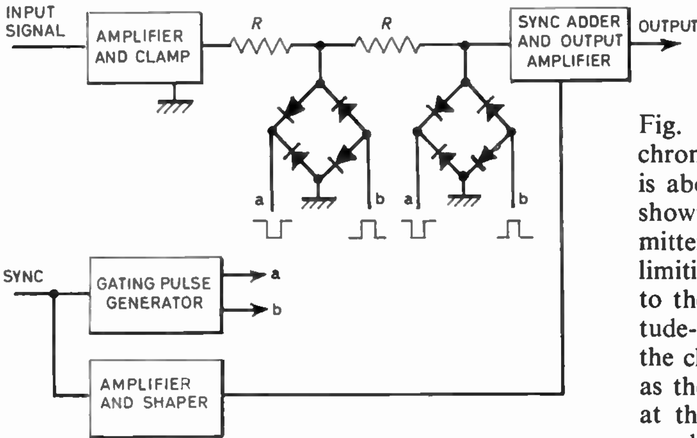


Fig. 6. Stabilizing amplifier.

of the mixer with a reference burst in a difference amplifier. Any discrepancies between the two bursts give an output which is added at the mixer output as a correction signal. This technique stabilizes both the amplitude and the position of the burst in one operation. The reference burst is produced by a unit called 'a black level generator' which gives out a signal consisting of synchronizing pulses with a burst, this source being correctly phased at the comparison point. The burst stabilizing circuit is arranged to ignore variations in amplitude of the burst of 0·25 dB in order to allow operational tolerances.

This deals with the case of synchronous sources but of course the mixer must handle non-synchronous sources and in these cases the local black level cannot be used for the reference. In the case of black-and-white non-synchronous sources the synchronizing pulses put back at the mixer output by the stabilizing amplifier are derived from the incoming waveform at a point prior to the fading action. Similarly the burst used for comparison by the burst stabilizing amplifier is also derived from the incoming signal at the mixer input. Unlike the synchronizing pulses the burst is not amplitude-stabilized to be exactly 0·3 V for non-synchronous sources, but is kept at the level it appears at the input to the mixer. Any severe discrepancy from 0·3 V indicates need of equalization and should be treated as a fault condition.

Another function of the stabilizing amplifier is to limit the signal in black-and-white working to 0·7 V peak-to-peak, a function which has been found to be very necessary with negative modulation and inter-carrier receivers, since an overload of the transmitter gives an output into the receivers' sound channel giving the characteristic buzz on sound. In black-and-white this has been cured by effectively limiting the output to exactly 0·7 V. With colour this is not so easy and an examination of the colour bars shown in Fig. 7 explains the reason. The excursion of the chrominance signal associated with yellow and cyan is above the white level for bars of 95% saturation, showing the extra excursion necessary at the transmitter if these colours are not to be distorted. Simple limiting does not work at sub-carrier frequency due to the fundamental considerations of passing amplitude-limited signals over a band-limited circuit. If the chrominance was limited in exactly the same way as the luminance then by the time the signal arrived at the transmitter over the band-limited circuit the amplitude would be excessive again. The stabilizing amplifier must still retain its low frequency, or luminance, limiting circuits but these must not affect the chrominance signal. A circuit tuned to 4·43 MHz placed in series with the limiting diodes prevents them from affecting the chrominance signals. Care must be taken with the limiting circuitry since the back-biased diodes, which is their state before overload occurs, are notorious for introducing differential phase distortion. This gives rise to the necessity for a very low-capacitance diode in this application. As a result of much consideration no limiting or gain variation occurs on the chrominance signal. Experience to-date has shown this to be satisfactory since all colour picture generating devices such as cameras and telecines have effective clippers operating at 0·7 V in their colour analysis outputs. Effective control at this point, together with balancing of the clipping points in each channel with respect to the

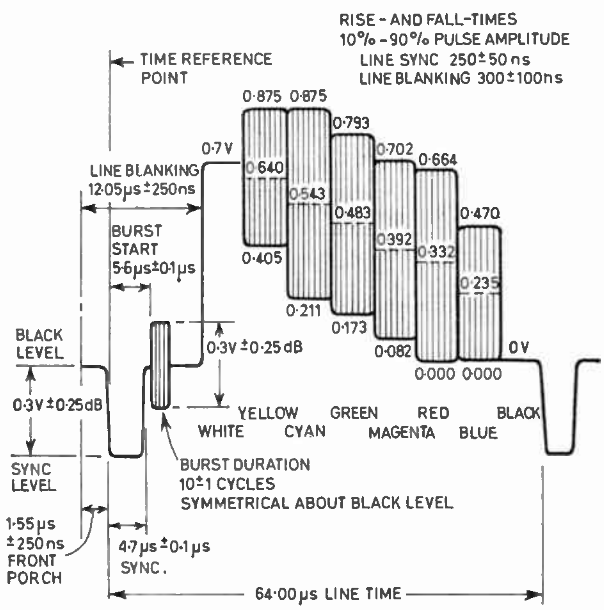


Fig. 7. 95% colour bars—line waveform.

others leads to a 'luminance only' output under overload conditions which is effectively dealt with by the luminance clipper.

6.2. *Special Effects*

Associated with all mixers are certain special-effects facilities which enable the change between sources to be made by means of wipes, also there is the ability to insert one picture within another which is known as inlay and overlay. With colour working these effects must be capable of being used without restriction and this requires that both sources being used are strictly synchronized for colour. Also the paths associated with both parts of the composite picture must be of identical electrical length and the luminance transition at the junction must be tailored to avoid the production of cross-colour effects.

One special effect used on black-and-white mixers, namely the black edging of captions, must be used with care on colour, due to the cross-colour effects produced on receivers. The inlaying of captions is preferable to the superimposition of captions where colour is concerned, and generally the colour contrast is such as to remove the necessity for black edging.

6.3. *Captions*

Captions are used to a great extent in a television service and if these have to be produced in colour then their cost would be greatly increased. A method of synthesizing a colour caption from a black-and-white caption has for some years been used in the B.B.C. and a control panel to operate such a unit, called a colour caption synthesizer, is placed on the mixer control desk. The choice of the colours for the letters and backgrounds has to be made with some care since it is easy to produce a caption which, although very effective in colour, is illegible in monochrome. The letter size of captions must also be chosen with care due to limitations of the colour system, as well as characteristics of the radio frequency modulation system used. The width of verticals associated with the letters of a caption must be sufficient to allow any colouring to be passed by the chrominance bandwidth available. Also the horizontals of the letters must be sufficiently wide to allow the averaging process of Delay PAL to leave them in a fit state for viewing. If these elementary considerations are not followed the resulting caption is severely degraded. Certain colour combinations, in particular, white letters on a red background and red letters on a white background, should be avoided because of the 'bleeding' effect produced at the transition between the letters and background. Such captions indicate without a shadow of doubt that the luminance/chrominance delay inequality permissible with colour systems employing the concept of luminance and chrominance is only about 40 ns total.

7. Monitoring

In order for the production staff to direct the programme and for the vision and lighting staff to control the output, colour pictures must be reproduced in the various operational areas. As a result of experience over the years it was decided to employ only picture monitors fed with the RGB signals and

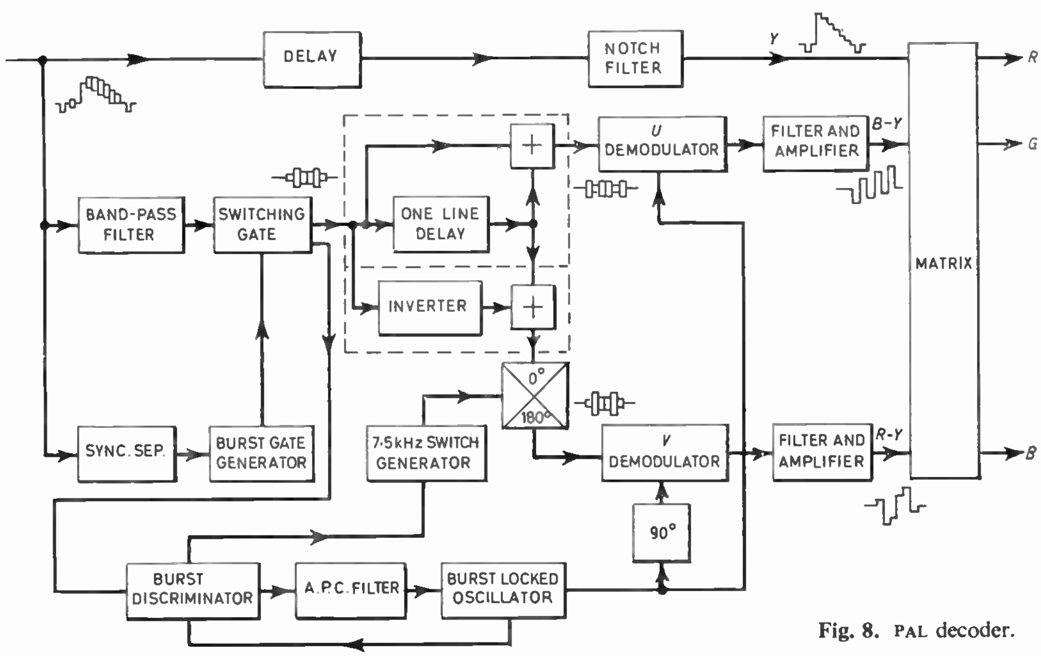


Fig. 8. PAL decoder.

to house the decoders associated with them in the apparatus room adjacent to the studio control room. With this technique any adjustments necessary to the decoder can be carried out without interfering with the monitor stack and also the decoder can be housed in a reasonably cool environment. The disadvantage of this procedure is the necessity to run four coaxial leads to each colour monitor. The picture monitoring can be carried out by using either one colour monitor for each source or only two colour monitors in each area, one for transmission and one for preview. The latter arrangement is very suitable for use with A/B type mixers as used in America but can lead to difficulties with 'knob-a-channel' mixers, as certain cases of difficulty can arise where picture matching is required. Consequently it would appear that a minimum of three monitors is probably the right number, as a colour monitor for each picture source would only be a liability rather than an asset due to the difficulty of keeping a large number of monitors accurately matched over a reasonable period of time.

8. PAL Decoder

Each colour monitor is driven from a professional decoder, which is capable of being switched between Simple PAL and Delay PAL as at one time it was thought that monitoring would use the Simple PAL mode of decoding to highlight any defect present on the signal. The block diagram of the decoder is shown in Fig. 8.

A composite video signal at the standard level is fed to the input of the decoder and is split three ways. The top route shown in the diagram consists of a delay network and a notch filter; the former is used to produce a delay in the luminance channel equal to that put in the chrominance channel due to filtering and demodulating, while the notch filter removes the chrominance sub-carrier signals from the luminance channel. These signals must be either removed or reduced appreciablly in amplitude otherwise they would give rise to a dot pattern on the colour picture as well as cause desaturation. The choice of width for the notch is a compromise between the elimination of the dot pattern caused by the colour sub-carrier and loss of definition in the colour picture. To remove all the dot pattern would require a notch having a bandwidth of 1·2 MHz below the sub-carrier, leaving a luminance bandwidth of only 3·2 MHz for the colour picture which tends to be inadequate. As the notch is narrowed the outer sidebands of the chrominance signal are left in the luminance signal giving rise to dot crawl on edges which on moving pictures can be annoying. In this decoder a choice of bandwidth of $\pm$ 500 kHz about the sub-carrier frequency has been made and to produce the optimum results the notch has some phase compensation applied to it.

From the incoming synchronizing pulses a burst gating pulse is derived which is applied to a switching gate. This switching gate accepts the output of a band-pass filter, handling those frequencies from 3·2 MHz to 5·5 MHz, and passes them during the active line period to the delay line separator. During the remaining period the chrominance signal, which consists of the burst, is applied to the burst discriminator where it controls the burst-locked oscillator producing the local sub-carrier used for demodulation. The result of this switching action is that the burst is not applied to the U and V demodulators, so no demodulated burst appears in the demodulated U and V signals. If the burst is allowed to be demodulated it then appears in the R, G, B signals where clamps in the picture monitor can clamp on it. When this occurs the black level of the monitor is a function of burst amplitude.

The circuits in the dotted area form the delay line separator to produce the U and V sub-carrier signals to be applied to the respective demodulators. The action of this delay line and the associated circuitry can be understood by examining the equations for the PAL system on two successive lines. These equations are as follows:

$$\text{Line } n: \quad E_U \cos \omega t + E_V \sin \omega t$$

$$\text{Line } n+1: \quad E_U \cos \omega t - E_V \sin \omega t$$

By means of the delay line these two signals can be made time coincident at the two adders. In the case where the signal on line $(n+1)$ is applied directly to the adder with the delayed signal from line n the term $2E_U \cos \omega t$ will appear at the output. When the signal on line $(n+1)$ is inverted before being applied to the adder the term $2E_V \sin \omega t$ results.

These two outputs are then applied to the respective demodulators, directly in the case of the U demodulator and via a 180° switch for the V demodulator. Filters placed in the outputs of the two demodulators, which are ring type, shape the chrominance signals and scale them to produce R$-$Y and B$-$Y. These two signals, together with the luminance signal, are applied to a matrix to derive the R, G, B signals at the standard level.

9. Signal Distribution

The signal has now progressed as far as the output of one studio and is then passed over the routing system within the studio complex before being routed to the transmitter. Important units in the routing system are switching matrices and for handling colour particular care must be taken in their design to ensure that their differential phase and gain distortion is very small. An additional requirement is that the crosstalk between routes must be very small and a figure of 70 dB is the design figure. If such a figure

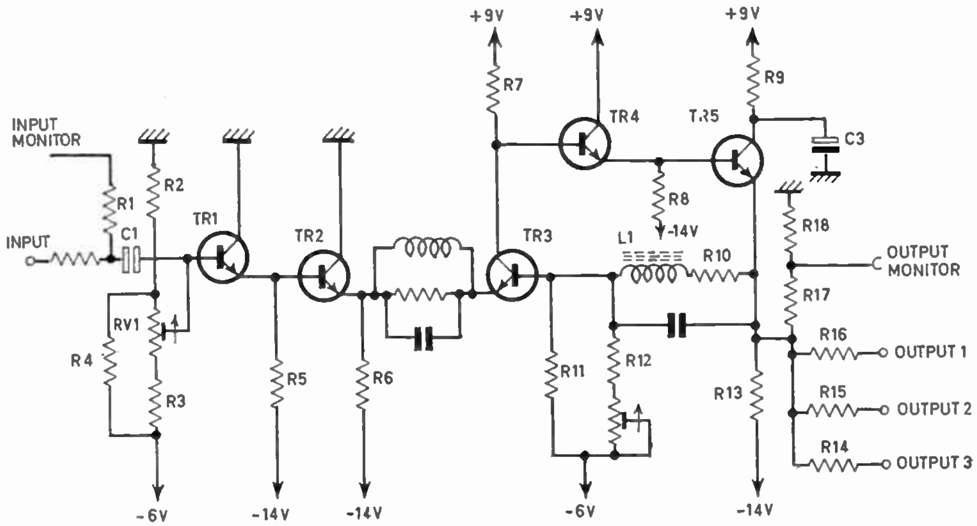


Fig. 9. Video distribution amplifier.

is not maintained, colour errors can result if the two sources are synchronous or, if they are non-synchronous, a colour pattern can result. When crosstalk figures of 60 dB are encountered the performance is marginal. The required performance is obtained by paying particular attention to screening and signal earthing in the design and layout of the switching matrices.

The routes connecting a studio to the telecine area, the video tape area, and the central apparatus room must be designed to have a very high performance. Such a route usually consists of an equalizer, an amplifier and a distribution amplifier. A route made up using such equipment has a chrominance/luminance ratio of $\pm 0{\cdot}1$ dB, a chrominance-luminance delay of ± 5 ns, a differential gain of 1% and a differential phase of 0·25°.

Before a signal leaves a studio complex it can traverse up to 30 pieces of equipment which will include two mixers. The performance required for such a route is chrominance/luminance ratio $\pm 0{\cdot}5$ dB, chrominance-luminance delay ± 20 ns, differential gain 5%, and differential phase 2° typical, with a 'worst case' figure of 4°. This 'worst case' figure is

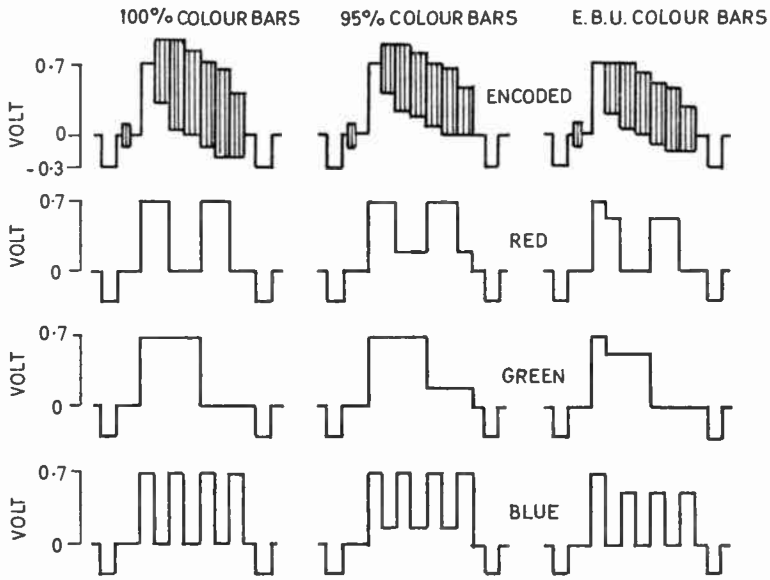


Fig. 10. Three versions of colour bars.

obtained with the C.C.I.R. signal having three blacks and three whites. To obtain these figures the distortion in each individual piece of equipment can only be very small.

One item which is used in large numbers in all signal distribution is the video distribution amplifier. Such an amplifier accepts the standard level signal and delivers at least three outputs at the standard level. Although the standard level in monochrome is 1 V with colour this increases to 1·23 V for saturated colours such as yellow and cyan. In order to handle this signal and also to cope with the changes in average picture level and bounce due to cutting between scenes a distribution amplifier must handle a signal of 3·2 V. The problem in the design of these amplifiers is one of maintaining a low differential phase distortion and this is achieved by having a basic design with very small inherent differential phase distortion[6] and then applying sufficient negative feedback which is really negative at the sub-carrier frequency. The differential phase performance of such an amplifier is 0·15° at 4·43 MHz, and the circuit is shown in Fig. 9. Some distribution amplifiers have a gain of 6 dB to allow equalization of short cable routes within a studio apparatus room.

The amplifiers used to equalize the routes between areas of the studio complex have a gain of 15 dB but only have one output. Their performance is similar in all respects to that of the video distribution amplifier.

10. Line-up of Equipment

10.1. *Colour Bars and Coder Line-up*

The addition of colour brings additional daily problems for the operational staff of a studio complex. These problems are generally associated with the line-up of the additional pieces of equipment; the first item is the coding equipment and this must be accurately lined up to obtain the required performance. The signal used to line up a coder is the colour bar signal of which there are three versions, but in this application only the 100% amplitude, 100% saturation version is used. Figure 10 shows for each version the composite signal, as well as the red, green and blue components. Examination of the components of each set shows that the 100% amplitude 100% saturation is the generic set and the other two versions are derived from it by taking 75% of each bar amplitude and adding a pedestal. In the case of the 95% bars this pedestal is added to all colours except black, whereas for E.B.U. bars the pedestal is added only to the white bar. Such a set of signals is generated by a colour bar generator and distributed to all coders in the studio apparatus room. The terminations are such that the level of each bar is, for the white bar, exactly 0·7 V; this equality being checked at the coder termination using a high impedance difference amplifier. Since each colour of the composite coded colour bar signal has a defined value of luminance and chrominance amplitude as well as sub-carrier phase angle (see Appendix), the output from a coder can easily be checked. From this it would seem that the coder line-up consists of

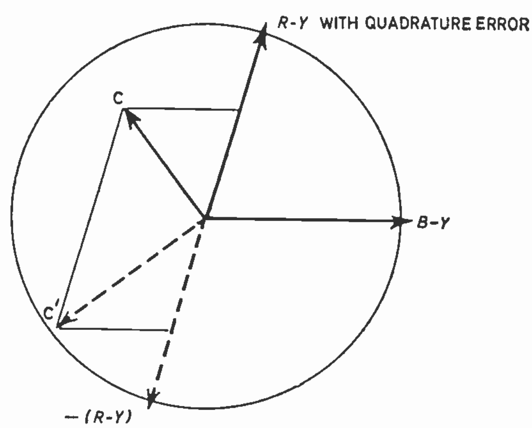


Fig. 11. The effect of a quadrature error in a PAL coder.

measuring the amplitudes of the luminance and chrominance signals and the phase angles of the chrominance associated with each bar. These measurements can be simplified for the chrominance signal to measuring the values of sub-carrier amplitude for the *U* and *V* signals and setting the angle between them to be 90°. Due to the switching of the *V*-axis in the PAL system even the measurement of the 90° angle is not necessary thereby enabling all coder line-up to be carried out using only an oscilloscope.

The measurement, or adjustment, of the 90° between the *U* and *V* axis is achieved by triggering the oscilloscope displaying the coder output waveform so that two adjacent lines overlap. In this condition any error in the 90° shows up as an amplitude twitter on the chrominance signal of each bar and is very easily detected. The mechanism by which this twitter arises is illustrated in Fig. 11; the vector **C** is the result of adding a component of B − Y to a component of R − Y, the two component vectors not being at right angles. When the R − Y component is switched through 180° it takes the position shown dashed in Fig. 11 and when this is added to the unswitched B − Y component the result is the vector **C**′, which is of a different length to vector **C** when the angle between the two components is not a right angle. Only when the two component vectors are at right angles will vector **C** equal **C**′. When the chrominance signals on two lines are superimposed as they can be on an oscilloscope the vectors **C** and **C**′ overlay each other and the *V*-axis switching action gives rise to the twitter. When the angle is set correctly the twitter disappears and in this condition experience has shown that the value of 90° is correct within $\pm 0{\cdot}5$ °. Therefore, by utilizing this effect, all setting-up of PAL coding equipment can be carried out using only an oscilloscope. With the N.T.S.C. system the vectorscope was a necessary instrument for coder line-up to adjust the 90° between axes, but when using the PAL system the vectorscope is an unnecessary as well as sometimes a misleading instrument. It is misleading because it cannot be provided with a sufficiently accurate means of adjustment prior to being used as a measuring instrument.

Having set the 90° angle between *U* and *V* axes to be correct and checked that the three colour bar components have the correct and equal levels for white, the carrier balances at black-and-white are then checked but should need no adjustment. The luminance is checked to see that there is a level of 0·7 V on the white bar and the value of *U* is checked on the yellow bar and the value of *V* on the red bar. These amplitude measurements are done using a calibrating waveform in the same manner as the 1 V standard level is measured.

10.2. *Decoder Line-up*

The line-up of the decoders, whether they be integral in monitors or separate on racks is basically the reverse process and uses colour bars of the 100% version. The fundamental requirement of the decoder is to deliver signals of identical amplitude on white from each of its outputs. The ability to switch between Simple and Delay PAL enables the angle between the demodulators to be accurately checked, as any departure from quadrature appears as the Hanover bar pattern in the Simple PAL position. When this adjustment has been made the decoder can then be worked in the Delay mode and the luminance and saturation controls adjusted to give the correct amplitude and shape for the blue bar output. Adjustment of the *V*-axis gain gives a correct red bar output.

10.3. *Picture Monitor Line-up*

The other items of colour equipment requiring special line-up are the colour picture monitors. A grill pattern is capable of being applied to all colour picture monitors to enable convergence to be checked and adjusted if necessary. The exactitude of the convergence of colour monitors is, at present, very dependent upon the mains voltage and the ambient temperature as well as the number of convergence controls available. As time progresses it is hoped that the convergence of picture monitors will be superior to that of commercial receivers.

Having adjusted the convergence the grey scale must be tracked and adjusted to be illuminant D. A step wedge varying from black to white is applied to each monitor and by means of the background, slope, and individual colour gain controls the grey scale is adjusted to be illuminant D at three points in the contrast range corresponding to black, grey and white. To assist in this operation a reference calibrated source of illuminant D is provided adjacent to the monitor stack. Again the stability of the grey scale is very dependent on temperature and should be improved in due course. Devices are being constructed to allow the operator to compare more easily the reference illuminant D with the grey scale on the monitor. The setting of the monitor brightness and contrast is achieved using the well-known picture line-up generating equipment (P.l.u.g.e.) signal[7] which is just as powerful in colour as monochrome.

11. Acknowledgments

The author is grateful for the assistance given by Mr. G. D. Roe of the B.B.C. Designs Department in the preparation of the diagrams used in the paper.

Finally, the author wishes to thank the Director of Engineering of the British Broadcasting Corporation for permission to publish this paper.

12. References

1. Joint I.E.E.I.E.R.E. Colloquium on Colour Television Cameras, March 1967.
2. James, I. J. P. and Karwowski, W. A., British Patent No. 1,046,777.
3. Bruch, W., 'PAL Selected Papers', March 1965; 'Selected Papers II, PAL a variant of the N.T.S.C. colour television system', June 1966. Both published by *Telefunken-Zeitung*, Telefunken A.G., Berlin.
4. Rogers, B. J., 'The PAL colour television system', *The Radio and Electronic Engineer*, 33, No. 3, p. 147, March 1967.
5. Bliss, J. L., Miller, I. D. B., and Allen, C. J., 'Picture source synchronization: the Natlock system', *EBU Review*, No. 107-A, p. 2, February 1968.
6. Broughton, D. C. and Parker, F. G., 'Improvements in and relating to transistor amplifiers', British Patent No. 977, 921.
7. Anstey, H. G. and Ward, P., 'Operational control of television studio picture quality', 'Television Engineering', I.E.E. Conference Report Series No. 5, pp. 286–92, 1963. (I.E.E. Paper No. 3976E, June 1963.)

13. Appendix

Colour Bar Parameters

The tables below give the amplitude in volts of the luminance signal measured from the black level, the peak-to-peak value of the *U* and *V* sub-carrier components and the peak-to-peak value of the total chrominance sub-carrier component for each colour. The angle of each colour is given relative to the (B−Y) axis. The angles given for line *n* correspond to the odd lines of the 1st and 2nd fields and the even lines of the 3rd and 4th fields. The angles given for line $n+1$ correspond to the even lines of the 1st and 2nd fields and the odd lines of the 3rd and 4th fields.

Table 1
100% colour bars

Colour	Luminance	U	V	Total chrominance	Angle in degrees line n	Angle in degrees line $n+1$
White	0·700	0	0	0	—	—
Yellow	0·620	0·612	0·140	0·627	167	193
Cyan	0·491	0·206	0·861	0·885	283·5	76·5
Green	0·411	0·405	0·721	0·827	240·5	119·5
Magenta	0·289	0·405	0·721	0·827	60·5	299·5
Red	0·209	0·206	0·861	0·885	103·5	256·5
Blue	0·080	0·612	0·140	0·627	347	13·0
Burst	0			0·300	135	225

Table 2
95% colour bars

Colour	Luminance	U	V	Total chrominance	Angle in degrees line n	Angle in degrees line $n+1$
White	0·700	0	0	0	—	—
Yellow	0·640	0·459	0·105	0·470	167	193
Cyan	0·543	0·155	0·646	0·664	283·5	76·5
Green	0·483	0·304	0·541	0·620	240·5	119·5
Magenta	0·392	0·304	0·541	0·620	60·5	299·5
Red	0·332	0·155	0·646	0·664	103·5	256·5
Blue	0·235	0·459	0·105	0·470	347	13·0
Burst	0			0·300	135	225

Table 3
EBU colour bars

Colour	Luminance	U	V	Total chrominance	Angle in degrees line n	Angle in degrees line $n+1$
White	0·700	0	0	0	—	—
Yellow	0·465	0·459	0·105	0·470	167	193
Cyan	0·368	0·155	0·646	0·664	283·5	76·5
Green	0·308	0·304	0·541	0·620	240·5	119·5
Magenta	0·217	0·304	0·541	0·620	60·5	299·5
Red	0·157	0·155	0·646	0·664	103·5	256·5
Blue	0·060	0·459	0·105	0·470	347	13·0
Burst	0			0·300	135	225

Manuscript first received by the Institution on 18th April 1968 and in final form on 1st July 1968.
(Paper No. 1215/Com. 6.)

Of Current Interest . . .

Paper Transistors

At the recent International Conference on Solid-State Devices held in Manchester, an American research scientist, Dr. T. P. Brody, of the Westinghouse Electric Corporation, described a process which allows the 'printing' of transistors on such common throw-away materials as paper, plastic kitchen wrap and aluminium foil. The new process stencils the transistors on these bases, or substrates, with vapours of metals and glass. The development of these devices is still in its early stages and a commercial product is still some years away.

The transistors are about the size of the crossbar on this printed letter 'T' and much thinner than the layer of ink used to print it. They can be bent, twisted and coiled—a property unique among electronic devices. This flexibility offers the possibility of 'printing' them on data cards, in books, or elsewhere, and suggests their mass production in the form of continuous rolls or strips.

Dr. Brody said that thin-film transistors on paper, foil or plastic could lead to inexpensive applications such as toys, hobby kits, novelties, teaching aids and other expendable items and in fact they appeared to be useful for almost any device that does not have to operate at high temperatures or be especially rugged, or does not involve high power or very high frequency.

In a typical paper transistor tellurium is the semi-conductor element and other parts are of gold, glass (silicon monoxide) and aluminium. The flexible transistors are made by evaporating their component materials, one after another, through metal masks, or stencils. The transistors are less than one hundred-thousandth of an inch thick, built up by four vapour deposits.

The process departs completely from the idea that thin-film devices can only be fabricated on expensive, ultra-smooth and rigid insulating materials such as sapphire, quartz and glass. In explaining this, Dr. Brody pointed out that while a rough-textured paper may look like a mountain range through the eye of a microscope the tiny molecules of a thin film 'see' its peaks and valleys only as a gently rolling plane. One thing is just about as rough as another on the atomic scale of things.

Also, it departs from the usual in-vacuum fabrication of thin-film devices a single layer at a time, with the vacuum chamber being pumped down after each layer is deposited. Instead, with a single pump-down a roll of paper or foil is passed through a 'printer' one full frame at a time. At present each frame is about the size of a postage stamp and has more than 600 transistors on it. Rolls of more than 13 000 transistors have been fabricated, but there appears to be no inherent limit to the length of the roll. These transistors can be flexed repeatedly without damage. They can even be cut in half and each part will continue to function.

The Westinghouse scientists have obtained 100% yields of working devices on many arrays of 100 transistors. They have been operated, without encapsulation, for more than 1000 hours with no measurable loss in performance.

In addition to single transistors, the process is reported to have been adaptable to the 'printing' of complete electronic integrated circuits on a continuous tape. The entire process lends itself to automation, with the printing, testing and encapsulating of components or circuits as one continuous operation.

Among the applications foreseen for paper-, plastic- and foil-based devices, Dr. Brody cited: circuits on data cards, documents and credit cards to facilitate electronic processing; teaching manuals with working circuits printed in them; pads of tear-off circuits for laboratory and home workshop experiments; and electronic implants and sensors for medical and biological use.

Satellite Earth Stations for Jamaica, Hong Kong and Kenya

A commercial satellite earth station is planned to be operational in Jamaica in 1970: announcements of this project were made on 8th August in Kingston, Jamaica, by the Hon. N. C. Lewis, Minister of Communications and Works, and in London by Colonel Donald McMillan, Chairman of the Cable and Wireless Group. The station will be the fourth to be owned and operated by the Group. Colonel McMillan said that the Group was also planning for other satellite earth stations in the Caribbean.

The Jamaica project will cost more than £2 million, and when built the earth station will relieve the pressure of considerable traffic expansion on the Jamaica-Florida coaxial telephone cable which has been in service since 1963. It will be constructed to withstand hurricane-force winds; the aerial structure will be a 90/100 ft dish.

The Group's first satellite earth station has been operational on Ascension Island for the U.S. *Apollo* project since 1966,† and other earth stations are being constructed at Hong Kong and Bahrain to be in service early in 1969. The Hong Kong station will transmit and receive via a Pacific satellite. Some 3000 tons of rock have been blasted to level the Stanley Peninsula site for this station and in readiness for a second Hong Kong earth station with a westward facing aerial structure.

Cable and Wireless is also engaged as technical consultant for the East Africa earth stations and a statement recently issued in Nairobi by the Board of East African External Telecommunications Company Ltd. states that for the construction of a satellite earth station at Mount Margaret in the Rift Valley, Kenya, the tender of the Marconi Company Ltd. has been accepted.

The Group is responsible for the external telecommunication services for fourteen territories in the Caribbean. It operates in 53 countries throughout the World and is actively pursuing a progressive training policy for locally-engaged staff and where earth station projects are planned or in progress there are career opportunities for local staff in satellite systems and, in certain countries, for computer working. Over 90% of the 10 000 staff are nationals of the countries where the Group operates.

† 'Satellite communications ground station on Ascension Island', *The Radio and Electronic Engineer*, **31**, No. 3, p. 159, March 1966.

U.D.C. 621.372.853.1

An Approximate Calculation of the Phase Constant for a Rectangular Waveguide Containing a Dielectric

By
Professor
O. G. VENDIK, D.Sc.†

Summary: The phase constant of a rectangular waveguide containing a longitudinal dielectric slab of arbitrary cross-section has been calculated using a system of orthonormal waveguide functions in which the curl and potential fields are separated.

1. Introduction

There are several methods for calculating the phase constant for a waveguide containing a dielectric slab which is homogeneous along the waveguide longitudinal axis but arbitrary in its cross-section.[1-4] The simplest one is the method of perturbation theory for the case when the inner electric field is calculated by solving Laplace's equation in the so-called quasi-static approximation. This approximation is applicable if the slab is considered to be in a homogeneous field which is not disturbed by the presence of the waveguide walls. The application of Laplace's equation for finding the inner field in the dielectric slab can be extended to the case of a slab of arbitrary form and dimensions if the real distribution of propagating waves and the influence of the waveguide walls can be taken into account. This is possible using the waveguide potential fields. In a previous paper[5] the orthogonal system of the waves for a rectangular waveguide, in which solenoidal and potential fields were separated, was obtained. Use of this system enables one to calculate the phase constant for arbitrary dielectric slab, taking into account only the dominant propagating modes, because the correct value of the inner field in the slab is expressed through the function of merely the potential-type field. In this case the problem does not require the solution of a characteristic equation of high degree and this is the advantage of this method in comparison with the known variational methods.[2, 4]

2. Theoretical Analysis

We will represent the effect of the dielectric placed in the waveguide by its associated polarization. If the dielectric permittivity tensor is $\varepsilon_{\alpha,\alpha'}$ the polarization may be given as a function of the inner field in the dielectric:

$$\pi^{(\alpha,\, e)}(x, y, z) = \varepsilon_0 \cdot (\varepsilon'_{\alpha,\,\alpha'} - \delta_{\alpha,\,\alpha'}) \cdot \prod\nolimits^{(\alpha',\, e)}(x, y, z) \qquad \ldots\ldots(1)$$

The notation used here is the same as that given in Ref. 5, and the index e signifies that we are considering the electric field $\Pi^{(e)}$ and the electric polarization $\pi^{(e)}$.

Using eqn. (1) and the expressions (14)–(16) from Ref. 5 (see Appendix) we obtain the main equations

$$\begin{aligned} C^{(\zeta)}_{\mu\nu}(\gamma^2_{\mu\nu} - \gamma^2) + G_\zeta \sum_{\mu'\nu'} \sum_{\zeta'=1}^{6} C^{(\zeta')}_{\mu'\nu'} \times \\ \times \sum_{\alpha,\,\alpha'=1}^{3} (\varepsilon'_{\alpha,\,\alpha'} - \delta_{\alpha,\,\alpha'}) \times \\ \times U^{(e,\,\alpha,\,\zeta)*}_{\mu\nu}\, U^{(e,\,\alpha',\,\zeta')}_{\mu'\nu'} \times \\ \times \iint_{\Delta S} F^{(e,\,\alpha)}_{\mu\nu}(x, y) F^{(e,\,\alpha')}_{\mu'\nu'}(x, y)\, dx\, dy = 0 \end{aligned} \qquad \ldots\ldots(2)$$

where $F^{(e,\,\alpha)}_{\mu\nu}(x, y)$ are the membrane functions

$U^{(e,\,\alpha,\,\zeta)}_{\mu\nu}$ are the sets of the coefficients which determine the relation between the different components of the fields in the waveguide,

μ and ν are the integers characterizing the transverse field distribution in the empty waveguide, and

ζ is a factor which determines the mode; ζ takes values from 1 to 4 corresponding to the curl modes; $\zeta = 5$ and 6 correspond to the electric and magnetic potential fields.

The numbers G_ζ are different for the different values of ζ.

$$\left.\begin{aligned} G_{1,\,3} &= k\left[k + \sqrt{\left(\frac{\mu\pi}{a_x}\right)^2 + \left(\frac{\nu\pi}{a_y}\right)^2 + \gamma^2}\,\right] \\ G_{2,\,4} &= k\left[k - \sqrt{\left(\frac{\mu\pi}{a_x}\right)^2 + \left(\frac{\nu\pi}{a_y}\right)^2 + \gamma^2}\,\right] \\ G_{5,\,6} &= \gamma^2_{\mu\nu} - \gamma^2 \end{aligned}\right\} \qquad \ldots\ldots(3)$$

where a_x, a_y are the waveguide dimensions,

$\gamma_{\mu\nu}$ is the phase constant of the empty waveguide,

γ is the phase constant of the dielectric-loaded waveguide which is to be found.

† Special Physics Department, V. I. Ulianov (Lenin) Electrical Engineering Institute, Leningrad, P.22.

The equations for the amplitudes of the solenoidal and potential fields can be considered separately

$$C_{\mu\nu}^{(\xi)}(\gamma_{\mu\nu}^2-\gamma^2)+G_\xi \sum_{\mu',\nu'} \sum_{\xi'=1}^{4} \times \\ \times [A_{\mu\nu,\mu'\nu'}^{(\xi,\xi')} C_{\mu'\nu'}^{(\xi')}+A_{\mu\nu,\mu'\nu'}^{(\xi,\lambda)} \cdot C_{\mu',\nu'}^{(\lambda)}]=0 \quad \ldots\ldots(4)$$

$$C_{\mu\nu}^{(\lambda)}+\sum_{\mu'\nu'} A_{\mu\nu,\mu'\nu'}^{(\lambda,\lambda)} C_{\mu'\nu'}^{(\lambda)}=-\sum_{\mu'\nu'} \sum_{\xi'=1}^{4} A_{\mu\nu,\mu'\nu'}^{(\lambda,\xi')} C_{\mu'\nu'}^{(\xi')} \quad \ldots\ldots(5)$$

where

$$A_{\mu\nu,\mu'\nu'}^{(\xi,\xi')}=\sum_{\alpha,\alpha'=1}^{3}(\varepsilon'_{\alpha,\alpha'}-\delta_{\alpha,\alpha'})U_{\mu\nu}^{(e,\alpha,\xi)*}U_{\mu\nu}^{(e,\alpha',\xi')}\times \\ \times \iint_{\Delta S} F_{\mu\nu}^{(e,\alpha)}(x,y)F_{\mu',\nu'}^{(e,\alpha')}(x,y)\,dx\,dy \quad \ldots\ldots(6)$$

The indices ξ correspond to the curl waves, the index λ corresponds to the potential fields.

The system of linear equations (5) can be solved for $C_{\mu\nu}^{(\lambda)}$ taking into account the terms with $C_{\mu'\nu'}^{(\xi')}$ as the constant members of the system. We shall use the method of approximate solution which is known as the iteration method.[6]

Finding the solution of eqn. (5) in the form of an iteration series and substituting it into eqn. (4) we obtain:

$$C_{\mu\nu}^{(\xi)}(\gamma_{\mu\nu}^2-\gamma^2)+G_\xi \sum_{\mu'\nu'} \sum_{\xi'=1}^{4} \Gamma_{\mu\nu,\mu'\nu'}^{(\xi,\xi')} C_{\mu'\nu'}^{(\xi')}=0 \quad \ldots\ldots(7)$$

where

$$\Gamma_{\mu\nu,\mu'\nu'}^{(\xi,\xi')}=A_{\mu\nu,\mu'\nu'}^{(\xi,\xi')}-\sum_{\alpha,\beta}\frac{A_{\mu\nu,\alpha\beta}^{(\xi,\lambda)}A_{\alpha\beta,\mu'\nu'}^{(\lambda,\xi')}}{1+A_{\alpha\beta,\alpha\beta}^{(\lambda,\lambda)}}- \\ -\sum_{\alpha\beta}\sum_{\alpha'\beta'}\frac{A_{\mu\nu,\alpha\beta}^{(\xi,\lambda)}\cdot A_{\alpha\beta,\alpha'\beta'}^{(\lambda,\lambda)}\cdot A_{\alpha'\beta'\mu'\nu'}^{(\lambda,\xi')}}{(1+A_{\alpha\beta,\alpha\beta}^{(\lambda,\lambda)})(1+A_{\alpha'\beta',\alpha'\beta'}^{(\lambda,\lambda)})}-\ldots \quad \ldots\ldots(8)$$

A significant simplification of eqn. (8) can be made if the dielectric permittivity is scalar. In this case it is possible to suppose that

$$A_{\mu\nu,\mu\nu}^{(\lambda,\lambda)} \cong (\varepsilon'-1)\frac{\Delta S}{S} \quad \ldots\ldots(9)$$

where S and ΔS are the cross-section areas of the waveguide and the dielectric slab respectively.

Thus

$$\Gamma_{\mu\nu,\mu'\nu'}^{(\xi,\xi')}=\frac{\Delta S}{S}(\varepsilon'-1)\left\{{}^{(o)}F_{\mu\nu,\mu'\nu'}^{(\xi,\xi')}-\sum_{p=1}^{\infty}\frac{{}^{(p)}F_{\mu\nu,\mu'\nu'}^{(\xi,\xi')}}{\left(\frac{\Delta S}{S}+\frac{1}{\varepsilon'-1}\right)^p}\right\} \quad \ldots\ldots(10)$$

where

$${}^{(p)}F_{\mu\nu,\mu'\nu'}^{(\xi,\xi')}=\frac{S/\Delta S}{(\varepsilon'-1)^{p+1}}\underbrace{\sum_{\alpha\beta}\ldots\sum_{\alpha'\beta'}}_{p\ \text{times}}\times \\ \times\underbrace{A_{\mu\nu,\alpha\beta}^{(\xi,\lambda)}A_{\alpha\beta,\alpha''\beta''}^{(\lambda,\lambda)}\ldots A_{\alpha'''\beta''',\alpha'\beta'}^{(\lambda,\lambda)}A_{\alpha'\beta',\mu'\nu'}^{(\lambda,\xi')}}_{p+1\ \text{times}} \quad \ldots\ldots(11)$$

If the set of indices $\left(\frac{\xi}{\mu\nu}\right)$ is denoted by one suffix n the system (7) becomes

$$\left.\begin{array}{l} C_1(\gamma_1^2-\gamma^2+G_1\Gamma_{11})+C_2G_1\Gamma_{12}+\ldots+C_nG_1\Gamma_{1n}=0 \\ C_1G_2\Gamma_{21}+C_2(\gamma_2^2-\gamma^2+G_2\Gamma_{22})+\ldots+C_nG_2\Gamma_{2n}=0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ C_1G_n\Gamma_{n1}+C_2G_n\Gamma_{n2}+\ldots+C_n(\gamma_n^2-\gamma^2+G_n\Gamma_{nn})=0 \end{array}\right\} \quad \ldots\ldots(12)$$

One can obtain an approximate solution of this system merely by using some of the equations containing C_n with restricted value of n. The simplest solution can be obtained by using only two values of n, corresponding to the dominant mode of the considered waveguide. After equating the system determinant to zero we obtain the characteristic equation

$$\gamma^2-\gamma_1^2-2k^2\Gamma_{11}=0 \quad \ldots\ldots(13)$$

Taking four values of n leads to the fourth-order characteristic equation

$$(\gamma^2-\gamma_1^2-2k^2\Gamma_{11})(\gamma^2-\gamma_3^2-2k^2\Gamma_{33})-4k^4\Gamma_{13}^2=0 \quad \ldots\ldots(14)$$

An example of numerical calculation of the phase constant will be given later for dielectric slab configuration for which the exact solution is known.[3, 4] The waveguide configuration is shown on the figure. Only the dominant mode TE_{01} can propagate through

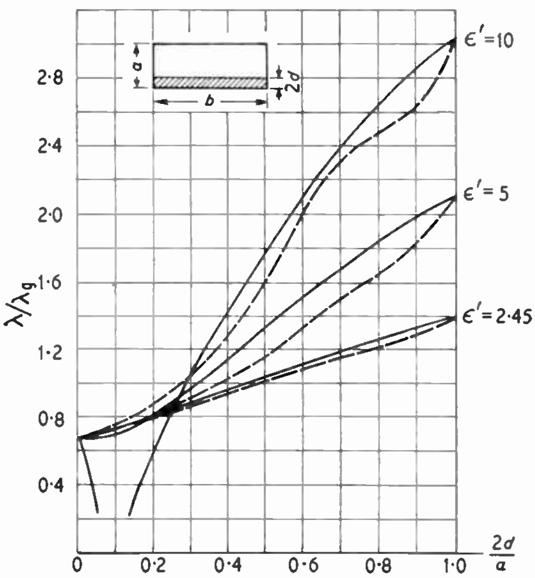


Fig. 1. Comparison of the approximate and exact solutions for the wavelength of the quasi-TE_{01} mode in a rectangular waveguide with a dielectric slab.

- - - - exact
——— approximate
$a = 0{\cdot}31\lambda \qquad b = 0{\cdot}72\lambda$

the waveguide. Only this mode will be used as a curl mode in our problem.

The integrals in eqn. (6) have to be calculated for the selected curl mode and for all electric potential modes. The result must be substituted into eqn. (11), where for simplicity we produce only one step of the iteration, i.e. $p = 1$. The summation over the indices of the potential field can be replaced by an integration, therefore we have the definite form for ${}^{(1)}F^{(1,\,1)}_{01,\,01}$. On substituting it into eqn. (10) and taking into account that p is supposed to be 1, eqn. (13) can be written in the following form

$$\left(\frac{\lambda}{\lambda_g}\right)^2 = 1-\left(\frac{\lambda}{2b}\right)^2+\frac{2d}{a}(\varepsilon'-1)\times \left[1-\frac{\dfrac{b}{8\pi d}\left\{1-\exp\left(\dfrac{-4\pi d'}{b}\right)\right\}}{\dfrac{2d}{a}+\dfrac{1}{\varepsilon'-1}}\right]$$

$$d' = \begin{cases} d & \text{for } d \leqslant \dfrac{a}{4} \\ \dfrac{a}{2}-d & \text{or } d \geqslant \dfrac{a}{4} \end{cases} \qquad \ldots\ldots(15)$$

This formula is obtained by using only one solenoidal mode, TE_{01}, and only one term from the iteration series for the potential fields. The numerical values given by eqn. (15) are shown in Fig. 1 together with the results of the exact analysis.[3, 4] The gap on the curve for $\varepsilon' = 10$ is the result of an unsatisfactory calculation of the potential field using only one iteration term. Using a greater number of iteration terms is straightforward using computer techniques. But in many cases when the dielectric slab is thin and convergence of the iteration series is bad, the usual quasi-static approximation is applicable.

It is very important to note that this technique is not restricted to a dielectric slab of rectangular cross-section but can be applied to any arbitrary cross-section.

3. Conclusions

The application of the method presented for the analysis of the wave propagation in a waveguide with arbitrary dielectric slab can be useful in some practical cases. This method supplements to a certain extent the known methods of the solution of the loaded waveguide problems.

In the theory, the calculations presented show that the main role in the description of the field configuration in loaded waveguides belongs to the potential fields and in this sense the separation of the potential and curl fields which has been made previously[5] is useful for practical calculations as well as for understanding the phenomena in a waveguide with complicated structure of the cross-section.

4. Acknowledgment

This paper was prepared during the tenure of a British Council Research Fellowship at the Department of Physics, University of Surrey, from October 1967 to April 1968.

The author expresses his sincere thanks to Dr. K. W. H. Foulds for interesting discussions of this problem.

5. References

1. Chatterjee, S. K. and Chatterjee, R., 'Dielectric loaded waveguides: A review of theoretical solutions', *The Radio and Electronic Engineer*, **30**, pp. 145–60, 195–205, 259–88, September, October, November 1965.
2. Collin, R. E., 'Field Theory of Guided Waves' (McGraw-Hill, New York, 1960).
3. Moreno, T., 'Microwave Transmission Design Data' (McGraw-Hill, New York, 1948).
4. Egorov, Yu. V., 'Partially-filled Rectangular Waveguides', (Sovetskoye Radio, Moscow, 1967).
5. Vendik, O. G., 'Excitation of waves in rectangular waveguide by the system of charges homogeneous along the waveguide axis', *Zh. Tekh. Fiz.*, **37**, pp. 1157–60, June 1967 (in Russian); English translation: *Soviet Physics—Technical Phys*, **12**, No. 6, pp. 835–8, 1967.
6. Faddeeva, V. N., 'Computational Methods of Linear Algebra' (Dover, London, 1959).

6. Appendix

The expressions from Ref. 5 are as follows:

Electric field in waveguide

$$\prod{}^{(\alpha,\,e)}(x, y, z) = \frac{1}{\sqrt{\omega\varepsilon}}\sum_{\zeta=1}^{6}\sum_{\mu\nu} C^{(\zeta)}_{\mu\nu}\,U^{(e,\,\alpha,\,\zeta)}_{\mu\nu}\,F^{(e,\,\alpha)}_{\mu\nu}(x, y)\,e^{-j\gamma z}$$

Amplitudes of the waveguide modes

$$C^{(\zeta)}_{\mu\nu} = -\frac{G_\zeta}{\gamma^2_{\mu\nu}-\gamma^2}\sum_{S=1}^{2}\sum_{\alpha=1}^{3} U^{(e,\,\alpha,\,\zeta)*}_{\mu\nu}\times \sqrt{\frac{\omega}{\varepsilon}}\iint_{\Delta S}\pi^{(e,\,\alpha)}(x, y)F^{(e,\,\alpha)}_{\mu\nu}(x, y)\,dx\,dy$$

The membrane functions $F^{(e,\,\alpha)}_{\mu\nu}(x, y)$ are expressed through the usual trigonometric functions. The table of functions $F^{(e,\,\alpha)}_{\mu\nu}(x, y)$ and coefficients $U^{(e,\,\alpha,\,\zeta)}_{\mu\nu}$ are given in another paper[5] by the author.

Manuscript first received by the Institution on 10th April 1968 and in final form on 17th July 1968. (Contribution No. 106/CC 23.)

U.D.C. 621.37/9.002:666.948

Processing Ceramics to give Suitable Substrate Characteristics

By

B. C. WATERFIELD†

Reprinted from the Proceedings of the Joint I.E.R.E.-I.E.E.-I.S.H.M. Conference on 'Thick Film Technology' held at Imperial College, London, on 8th–9th April 1968.

Summary: The selection of the substrate material is dependent on the application for which the deposited circuit is intended. The normal production specifications for substrates are discussed and a brief description given of the production methods available. The reasons for the impurity ratio in the ceramic must be considered relevant to the type of deposition that will be applied and the degree of adherence that is required; there is also a connection between impurity level and surface finish which will affect the choice of metallizing media. In terms of both cost and surface texture, 'as fired' ceramics are favoured. Any subsequent machining operations will tend to expose voids which can be clearly seen on roughness traces, though a general 'flattening' of crystal peaks will take place. If a higher degree of surface finish is obligatory, then great care is needed in the selection of ceramic with particular reference to the size of grain and alumina content. Finally, the factors affecting the costs of production substrates are considered in light of the aspects discussed.

1. Introduction

The ceramics industry is now equipped to supply substrates to the mechanical requirements of the user; various methods are employed with varying degrees of success and it is not the purpose of this paper to examine means of manufacturing. Rather it is the intention to concentrate on the physical properties of the substrate and to see how the variables in the processing of the ceramic material and in the constituents therein can affect the performance of the substrate in the hands of the user. The discussion throughout is limited to the group of ceramics with alumina contents greater than 85%.

2. Scope

Whilst alumina displays very excellent electrical properties its choice as a substrate material stems more from the thermal and mechanical characteristics; better heat transfer and a more robust vehicle for highly complex systems. However it has a number of disadvantages such as crystalline surface and a tendency to distort during sintering which requires a high degree of control. There are other variables, however, in the ceramic which can have an effect on the subsequent processing by the user particularly adherence of his films and the degree of accuracy in pattern due to surface texture irregularity. What are these variables and how are they produced; more important perhaps—how can they be detected and overcome?

3. Alnmina Snbstrate Makeup—Deliberate Impurities

Basically, the alumina ceramic is conceived as a majority content of calcined alumina crystals in a minor phase of oxidized glasses. Generally, the Al_2O_3 itself can be well controlled and the properties anticipated, the principal variation being the amount of NaO that is present after calcining. Conventional calcined high-purity alumina can contain up to 0·5% of soda, more desirable sources contain between 0·08% and 0·05%.

This apart the glass phase now emerges as the element of variation—variation within a given ceramic and of course from manufacturer to manufacturer. This phase will consist largely of silica with varying ratios of lime and magnesia. If the ceramic is to give reproducible properties then the control of these elements becomes very important. Only the most chemically pure sources can ensure that properties are maintained even if this means sacrificing some of the benefits in ease of processing that less pure materials might offer. For instance bentonite can, as a clay source, have certain 'in process' advantages but its high iron and titanium content can in reducing atmospheres cause considerable discolouration of the ceramic.

† Andermann & Ryder Limited, West Molesey, Surrey.

To examine in more detail the composition of high alumina ceramics we will examine some nominal analyses as found after sintering. (See Appendix.)

Obviously there will be some change of property with different alumina contents most noticeably dielectric constant, thermal conductivity and tensile strength, but alumina content alone cannot give a set of figures; that can only be done by taking into account the density and average grain size, and the type of secondary phase that is used.

It is therefore this factor of variation within ceramic bodies and between different sources of supply that has influenced the choice of 96% Al_2O_3 as the widely chosen minimum level and making the jump to 99% even more attractive if the risk of variables seems serious though in many cases it is the higher thermal conduction which is also influential.

3.1 *Incidental Impurities*

An alumina substrate on leaving the kiln is at its best. Sintering, carried out in an oxidizing atmosphere at temperatures in excess of 1450°C will have burnt out all organic matter, crystallites will have formed surrounded by the glass phase, the surface will be completely closed and above all it will be thoroughly clean. It cannot be overstressed that from this stage on there is every chance that impurities can be collected by the substrate due to the abrasive nature of its surface. Handling in any form whether by bare hand or tweezers will leave some residue. Throughout inspection and packaging and onwards every care must be taken to allow no metal pick up which will be easily converted to Fe_2O_3 during conductor firing. If therefore the user feels there is any risk of unclean substrates affecting his process then an air brake in excess of 950°C should be carried out.

In cases where grinding or lapping has to be carried out there is a great risk of the substrate collecting impurities. Coolants vary but are normally soda-based and in most cases one adds rust inhibitors such as sodium nitrate. Production grinding wheels are metal bonded and can work minute bronze or iron particles into the surface. The quality of the grinding can not only affect the texture of the surface but also its mechanical characteristics.

In addition to all this the surface now displays voids which will vary but could be up to 0·002 in across which will have collected swarf and will be cleaned with difficulty.

4. Ceramic/Metal Jointing

There is now a wide variety of methods of achieving a joint on to an alumina substrate all of which, to a greater or lesser extent, will be affected by the alumina content and therefore by the quality and composition of the glass phase. Apart from air-fired pastes both pure glassing and the sintered metal process all fall into the general category of thick films.

4.1 *Air-fired Pastes*

Here the adhesion of the metal takes place due to the adequate ratio of glass frit that is added to the metals. Since at the temperatures used in high alumina containing ceramics neither the alumina crystals nor the glassy phase will become mobile, the joint relies on the wetting properties of the frit. In a ceramic containing a large amount of silica this will tend to become mobile at the higher range of temperatures used and could, if not affecting the joint, make the metal layer unplatable. As in the matching of straight glazes the important factor is most likely to be the thermal coefficient of expansion with particular reference to the fusion temperature. Here too the size of the crystal will determine the volume of paint retained and the size of the exposed voids the degree of bridging that may take place. In any event the nature of the surface must affect line demarcation.

4.2 *Refractory or Sintered Metal Process*

With a trend toward greater strength of edge connectors to the lead frame continuing the use of brazing as an alternative to soldering must be considered. It also offers the user a range of temperatures and a surface to which welded or ball-bonded connections can be made and is often used for special packages.

Originally developed as a means of achieving high strength multiple seals for thermionic valves the aim was to achieve the maximum strength of joint possible, pulls being monitored at greater than 9000 lb/in^2 (630 kg/cm^2), half the tensile strength of the ceramic itself.

For this purpose molybdenum, due to its oxidation potential allowing control of the oxidation state in controlled atmosphere furnaces, has emerged as the principle constituent in a number of patented processes being used on its own as MoO_2 or more often accompanied by manganese, titanium and other reactive metals.

The fine-milled particles (1–5 μm) in a nitrocellulose suspension with acetone or highly volatile solvents are easily screened on to the ceramic and fired at temperatures in excess of 1300°C usually in a forming gas atmosphere.

Although there are many theories as to the mechanism of this type of joint there is no doubt that the composition and condition of the ceramic have a distinct effect on the strength of the joint or the ability to make a joint at all. The amount, mobility and even the viscous temperature of the glassy phase in the ceramic are all important. Very high alumina content will almost certainly require additions of some glass to the metallizing paint.

Another important variable is the grain size of the ceramic. Although optimum properties are achieved by a finely-knit structure a coarser grain may allow interchange between glassy phase and Mo layer and so achieve required strength.

It can be seen therefore, that whilst air-fired parts may be to some degree affected by the ceramic characteristics the refractory metal process may depend on them though it is generally agreed that the same degree of strength is not required. This could mean that the only safe way of taking advantage of this process is to buy-in the substrate already metallized with sintered nickel plating.

5. Surface Finish

It has been pointed out already that all ceramics develop special surface characteristics during sintering. The texture of the resultant finish will depend on the particle size of the basic alumina, the quantity and nature of glass present as the secondary phase and the peak temperature and pattern of the firing cycle.

Certainly there are chemical factors but there are also some mechanical considerations that have a bearing. It is not proposed to discuss the comparative virtues of producing the ceramic prior to firing, but pressing and film casting must be highlighted from one angle regarding the surface properties. At any stage the surface of the ceramic will reflect to some degree the surface of materials with which it is in contact. Hence in the pressing process the need is either to have punches of impeccable finish (in the region of 2 μm) and eliminate punch stiction, or to avoid contacting the surface at all by stamping from a film cast band of ceramic which, since it has been cast on to a glass surface of 1 μm finish, is as near perfect as one can expect to get.

The next variable occurs during firing where the above comments regarding contact surfaces again apply. The optimum method of firing is to stand the substrates on edge in which case both working surfaces would be open to the atmosphere. This, however, is rarely practicable on thin sections where excess bowing would occur. The alternative is 'flat' firing in which case the substrate will now in part reflect the surface quality of the batt. Stacking of substrates one upon the other would mean that they would tend to sinter together. This can be avoided by scattering crystals of corundum between layers, but the crystals however tend to penetrate the surface and even if removed by rumbling will leave a hollow impression.

Sintering is the vital part of the production process and here the majority of variables occur. The minority elements can encourage or retard grain growth thus affecting firing temperatures. The nearer to theoretical density the substrate can be the better the physical properties. Long firing schedules can achieve the same result as firing to a peak temperature; a long low fire will give less tendency to curl up during the re-forming phase of sintering and at the same time will give smaller grains and therefore smaller and a reduced quantity of voids.

Lower temperature firing might allow for electric heat sources though town gas is the more widely used. Here the ceramist is careful not to get an excess of reduction in the atmosphere and so tend to produce areas of incomplete binder removal showing up as dark shadows in the otherwise semi-translucent body and being of a carbonaceous nature. Again this is the one time in the process when minute traces of sulphur from the gas could penetrate into the mass and on cooling become trapped in minute quantities within the voids.

6. Conclusion

Though it might appear from these observations that the alumina substrate is highly susceptible to variation, and therefore a risk factor, these variables are known to ceramists who take endless pains to harness them. It does perhaps mean that the user is a little restricted in the breadth of choice he has for his substrates in that changing from supplier to supplier may bring drawbacks unless time can be taken to check for minor changes in the process. In fact it may still be too early to rely too much on price alone when the substrate source is being selected.

7. Appendix

Composition of high-alumina ceramics

Al_2O_3	SiO_2	CaO	MgO	Cr_2O_3
99·22	0·05	0·07	0·24	<0·01
95·44	3·11	0·20	0·59	0·09
93·09	3·18	2·04	0·70	0·01
Fe_2O_3	TiO_2	Mn_3O_4	P_2O_5	Na_2O
0·06	<0·01	<0·01	<0·01	0·11
0·11	0·01	0·02	0·06	0·18
0·22	0·01	<0·01	<0·01	0·39
K_2O	Li_2O	SnO_2		
0·03	<0·01	—		
0·05	<0·01	—		
0·03	<0·01	—		

Manuscript received by the Institution on 29th March 1968. (Contribution No. 107/CC 24).

U.D.C. 621.382.2:681.32

High-speed Computer Logic with Gunn-effect Devices

By

H. L. HARTNAGEL, Ph.D.†

AND

S. H. IZADPANAH, B.Sc.†

Summary: The application of the Gunn effect for fast logic circuits is discussed. In particular, several new devices, such as comparators, which rely on the possibility of domain-nucleating control are described. The ultimate time-constants of these new devices can be at least one order of magnitude better than the most advanced monolithic structures with conventional transistor and diode circuitry. Finally, two circuits, namely, an adder and a shift register, incorporating these new Gunn-effect devices are proposed.

1. Introduction

Development of computer-logic circuitry has reached the state where high speed of operation is possible. However, the frequency limitation due to the product of junction capacitance and characteristic impedance has now been reached, and a further increase in speed has to be obtained by paralleling logic elements as much as possible. At present the typical delay time for a gate is a few hundred picoseconds and cannot be decreased appreciably any more.

On the other hand, the recent development of semiconductor bulk-effect pulse generating and processing opens up the possibility of a much higher speed of operation. In particular, the domain properties of Gunn-effect diodes are suited for ultra-fast pulse electronics. For example, a Gunn diode can be made to produce pulses of only 100 ps duration when it is biased above a critical field, E_t. With these devices, the junction frequency limitation no longer exists and the frequency is determined by the diode length. The basic construction of these diodes consists of an n-GaAs crystal sandwiched between two ohmic contacts on opposite faces. To amplify or regenerate a pulse, the diode is biased just below the threshold field E_t, and an input pulse is superimposed on the biasing voltage. The field will momentarily be raised above threshold, causing the nucleation of a high field domain at the cathode. As soon as a domain is formed, the diode terminal current is reduced. This current increases to its original value when the domain reaches the anode and is discharged. The duration of the resulting current pulse is equal to the domain transit-time within the sample. The domain propagation velocity for GaAs is approximately 10^7 cm/s. Hence a crystal of 30 μm length would produce a pulse of 300 ps duration. The pulse repetition frequency (which depends on the triggering pulse frequency) could be made almost as high as the inverse of the domain transit-time. The current pulse is converted to a voltage pulse via a resistive load R_L.

The output pulse amplitude can be larger than the input pulse, i.e. one obtains gain. The maximum pulse gain is obtained if:

(i) The highest possible drop in diode current occurs when domain formation is achieved.

(ii) The value of R_L is made as high as possible, without interfering too much with the domain growth process.

Figure 1 shows the relation between the field

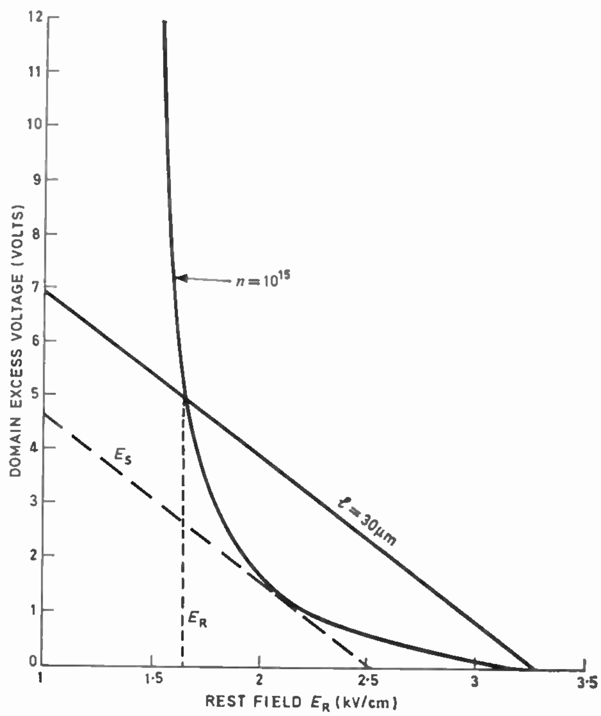


Fig. 1. Domain excess potential vs. rest field. (After Copeland.[1])

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

outside the domain, E_R, and the domain excess voltage, V_{ex}, for a particular carrier density.[1]

By selecting the diode length and carrier density one can obtain a minimum rest field E_R. This produces a maximum current drop, and condition (i) is satisfied.

In this way we have obtained voltage pulses of 4 V amplitude from a 33 μm-long diode working into a 7·5 Ω load resistance. These pulses can be used to trigger another diode in the successive stage as shown in Fig. 2. As the pulse produced by the first diode with positive biasing polarity is a negative pulse, the second diode which receives this pulse must be biased in opposite polarity, in order to obtain domain nucleation.

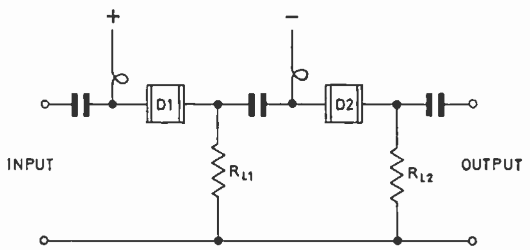


Fig. 2. Circuit of a two-stage amplifier.

A range of Gunn-effect logic devices has already been reported. The DOFIC[2] transforms analogue voltages into a succession of pulses, which can be employed for pulse code modulation by a suitable coding system. It relies on the effect that the distance a domain propagates can be controlled by an externally-applied voltage if the crystal has a trapezoidal shape.[3] A memory device utilizes the property that a single pulse can cause a Gunn oscillator to break into c.w. oscillations if the diode voltage and the external circuit are correctly chosen.[4] Logic gates have been obtained by the use of a third electrode along the length of the crystal.[4,5]

The authors have been working on further Gunn-effect logic devices such as an 'exclusive-OR' which is, for example, the basic element of an adder.[6,7] A number of other logic circuits can be developed, and some examples are discussed below.

2. The Comparator

Let us take a wafer of GaAs with a large contact on one face and a small contact on the opposite face. If a voltage V is applied to this diode with the small contact as cathode, there will be a high electric field at the cathode and a lower field at the anode. The magnitude of the high field is caused by the small size of the cathode.

By changing the cathode diameter, one can control the high-field value. Alternatively, one can make two small contacts separated by a gap g on the same face of a crystal, and a change in high field can be achieved by either connecting only one, or both of the contacts to the voltage source.

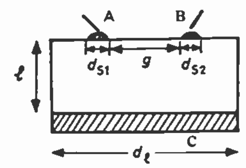


Fig. 3. A Gunn-effect comparator (triode).

A, B small electrodes as input terminals,
C large electrode (output terminal),
g distance between two small electrodes,
l sample length.

Figure 3 shows the configuration of such a device with two equal small electrodes A and B and a large electrode C.

It is appropriate to call this three-terminal device a 'Gunn-effect triode'. In the first case (a) one small electrode, say A, is connected to the negative and C to the positive biasing terminal; the electrode B is kept floating. In the second case (b), both small electrodes are connected together.

As an example, the potential distribution inside a crystal is shown in Figs. 4(a) and (b) for two cases. The field was computed by solving Laplace's equation on an *Atlas* computer using a successive over-relaxation method neglecting any space-charge layer along the metal-semiconductor interface. The bias voltage was assumed to be approximately $E_t l$, where l is the sample-length. The computation showed an area of relatively high field, i.e. $E > E_t$, near the connected small electrode. This is denoted in Fig. 4(a) by P_n. The field along the remaining length between P_n and the anode electrode was above the theoretical domain sustaining field E_s.

For the particular example given, the minimum field value occurring near the anode was 2·82 kV/cm. Taking suitable device parameters (ρ = 1 Ωcm and l = 30 μm), we find in fact the value of E_s from Fig. 1 to be approximately 2·5 kV/cm. Once a domain is formed it propagates therefore along the full length towards the anode.

In case (b), the computer showed a change in field magnitude at each point. The ratio of field change was sometimes larger than 3. This ratio is a direct

(a)

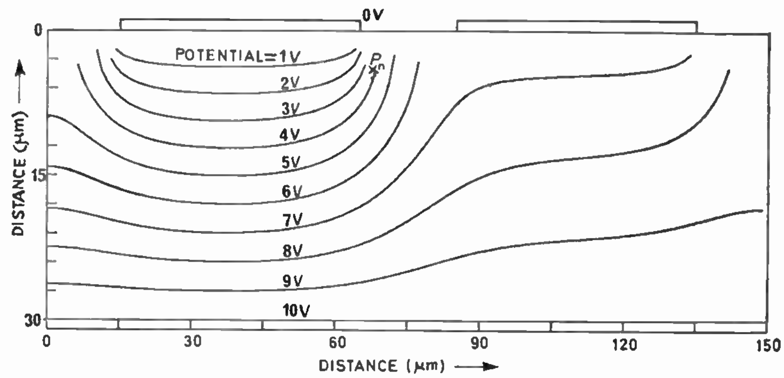
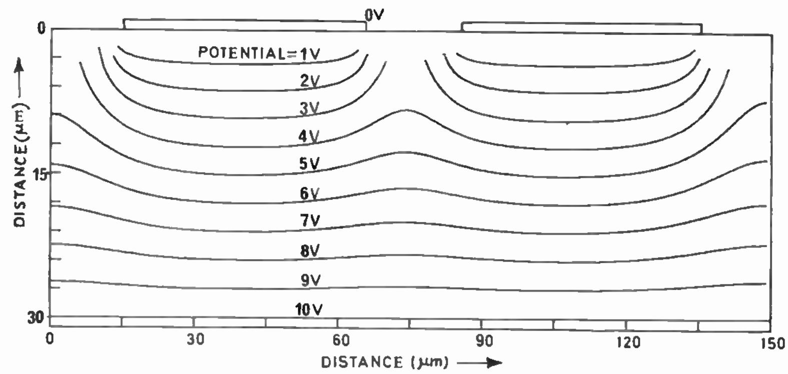


Fig. 4. (a) and (b) potential distribution of a Gunn-effect comparator.

P_n is the point of maximum field which is higher than the threshold field.

(b)



(c) Ratio of field variation for cases (a) and (b) at distances below small contacts.

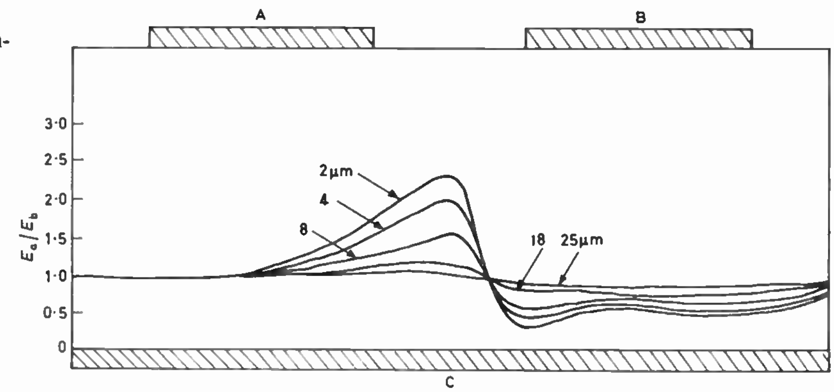


Table 1

Truth table of comparator

A	B	Output
0	0	0
1	0	1
0	1	1
1	1	0

function of g, d_{s1}, d_{s2}, d_l and l, where d_{s1} and d_{s2} are the diameters of the two small electrodes and d_l is the diameter of the large electrode. Especially, the field value at P_n was reduced.

The field variation for the two cases of this particular example is shown in Fig. 4(c). The pronounced reduction of the high field value in case (b) gives no chance for any domain formation as the maximum field is then below E_t. Hence, domain nucleation is obtained, if only one of the small electrodes is connected to the supply voltage; whereas no domains are formed if both electrodes A and B are connected to the same voltage. Therefore the device operates as a comparator whose truth-table is given in Table 1.

For a proper logic operation, one has to bias the comparator just below the threshold voltage and apply pulses with correct polarity and timing to the input electrodes A and B of Fig. 3.

To verify the theoretical results, some devices were constructed with epitaxial material supplied by Services Electronics Research Laboratory, Baldock, with a highly-conducting Si-doped substrate. In order to control the position and the size of the two small contacts (i.e. electrodes A and B of Fig. 3), a thin layer of tin was evaporated on one of the surfaces of the crystal through an appropriate masking. Then the samples were cut with an ultrasonic cutter. If a very small diameter of electrodes was desirable, a small pellet of tin was placed in addition on the evaporated electrodes and alloyed to the crystal around 500°C in a running forming-gas atmosphere in order to enable contacting wires to be applied. Otherwise a thick layer of gold was evaporated on the tin and alloyed afterwards. The device was housed in a suitable mount and placed in a coaxial cavity for preliminary measurements. If the device was giving the normal Gunn-effect oscillation, the experiments were then carried out with a resistive load. The applied bias voltage was always pulsed.

To examine case (a), a voltage was first applied between terminals A and C and the threshold voltage V_{t1} for domain formation was determined. Then the terminals B and C were subjected to a potential V_{t2}, which was just sufficient for producing Gunn oscillations. In most cases one found $V_{t2} \simeq V_{t1}$.

For case (b), the terminals A and B were connected together and a voltage was applied between this connection and C. Comparator action was considered to occur, when no output signals appeared in spite of the applied voltage being V_{t1} or V_{t2}, and when only an increase in applied voltage caused domain nucleation. Taking the threshold voltage for case (b), V_{tb}, we can now define a 'threshold ratio', K, by:

$$K = \frac{V_{tb}}{V_{t1}} \simeq \frac{V_{tb}}{V_{t2}}$$

Among the Gunn-effect triodes produced so far, only a few demonstrated clear comparator action. The failure of many devices can be explained by low crystal quality giving strong nucleating centres even outside the region P_n in Fig. 4(a).

A circuit giving another approach to the comparator is shown in Fig. 5. The device employs two ordinary Gunn diodes in parallel with a common resistive load R_L. The diode crystals are of equal length and equal cross-sectional surface and their contact areas are of equal size. The comparator action can be explained by using an appropriate resistive network, which of course, applies before any domain formation occurs. If one takes case (a) as described above, the threshold voltage V_t is related to the total applied voltage V_{B1} by

$$V_{B1} = V_t\left(1 + \frac{R_L}{R_O}\right) \qquad \ldots\ldots(1)$$

R_O is the small-field diode resistance.

For case (b), the biasing voltage V_{B2} is required for domain formation, i.e.

$$V_{B2} = V_t\left(1 + \frac{2R_L}{R_O}\right) \qquad \ldots\ldots(2)$$

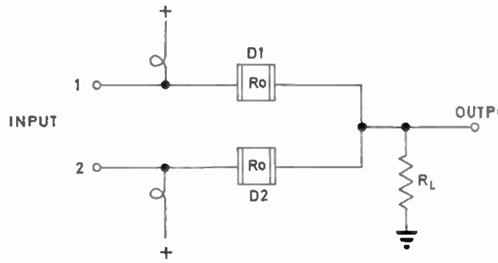


Fig. 5. Comparator circuit of two Gunn diodes in parallel.

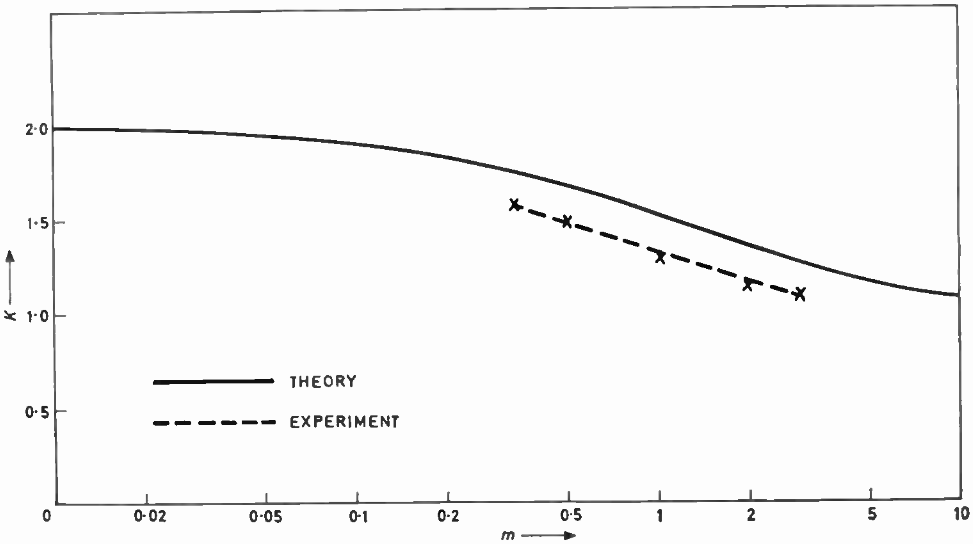


Fig. 6. Threshold ratio, K, as a function of $m = R_O/R_L$ for the circuit of Fig. 5.

One can see that

$$\frac{V_{B2}}{V_{B1}} = K > 1$$

The circuit will therefore act as a comparator with a threshold ratio K given as

$$K = 1 + \frac{1}{1+m} \qquad ......(3)$$

where

$$m = \frac{R_O}{R_L}$$

This shows that K varies from 1 to 2, as m takes values from infinity to zero. A reliable comparator is obtained when m is as small as possible. There will, however, be a smallest value for m for which a domain can still grow to a reasonable size, as a large R_L causes an increase in the domain growth time.

The comparator property of this circuit was verified experimentally and the results for diodes with l = 50 µm and $R_{O1} \simeq R_{O2} \simeq 60\Omega$ are compared with the numerical evaluation of equation (3) in Fig. 6. We have also inserted the three-terminal comparator in the same circuit as the two-parallel-diodes comparator.

One expects that the experimental threshold ratio should exceed then the theoretical values of Fig. 6. For several cases, this was in fact observed. The contribution of the three-electrode comparator together with the load resistor R_L caused the experimental threshold ratio to be further increased. The experimental results are given in Table 2. The increase in K caused by the Gunn-effect triode was often more pronounced for large R_L.

Table 2

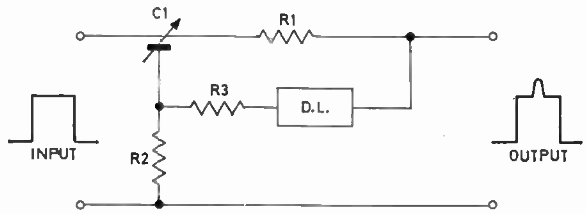
Representative experimental results of comparator

K_{th} = theoretical value of K

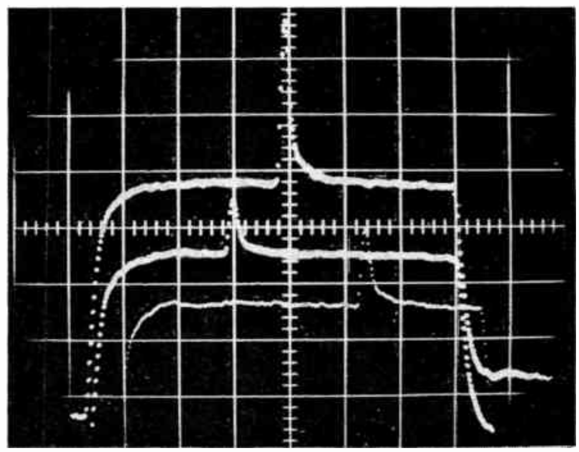
K_{exp} = experimental value of K

No.	l µm	d_{S1} µm	d_{S2} µm	d_t µm	g µm	m	K_{th}	K_{exp}
1	120	250	175	450	36·5	16	1·059	1·12
2	30	175	175	550	135	0·96	1·51	1·49
3	,,	,,	,,	,,	,,	0·45	1·69	1·7
4	,,	,,	,,	,,	,,	0·202	1·83	1·88
5	24	164	164	480	120	0·2	1·834	1·81
6	,,	,,	,,	,,	,,	0·06	1·945	2·14
7	,,	,,	,,	,,	,,	0·0357	1·965	2·16

For single-pulse operation, it is necessary to bias the sample below threshold. Our available material did not permit the application of a d.c. bias voltage



(a) Circuit for the production of a voltage spike, superimposed on a biasing pulse. DL = delay line.



(b) Typical waveforms produced by the circuit of Fig. 7(a) horizontal scale: 10 ns/div; vertical scale: 10 V/div.

Fig. 7.

as the heat produced would have destroyed the diodes. Therefore the devices were always supplied with a pulsed biasing voltage, and a triggering spike had to be superimposed on it. This was achieved as follows:

The bias pulse was also applied to a differentiating circuit and the resulting spike which had the same polarity as the original signal was superimposed on the pulse through a suitable delay. The circuit diagram and typical waveforms are shown in Fig. 7(a) and (b). The ratio of spike to bias pulse amplitude and their relative position and spike time constant could easily be controlled by C1, R1, R2, R3 and delay line DL in Fig. 7.

It is of great importance to consider the power dissipation P_d of the logic devices such as the comparator. This is:

$$P_d = \frac{V_a^2}{R_O} \qquad ......(4)$$

where V_a is the applied voltage. If $V_a \simeq V_t$, then

$$P_d = E_t^2 n l \mu_0 e S \qquad ......(5)$$

where n = carrier density, μ_0 = low-field mobility, e = electronic charge and S = diode cross-sectional area.

The power dissipation for a fixed nl is proportional to S. For example, a diode of $nl = 10^{12}$ 1/cm² with

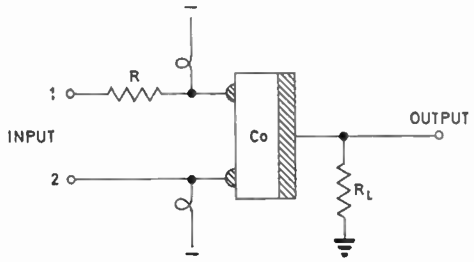


Fig. 8. A Gunn-effect inhibitor.

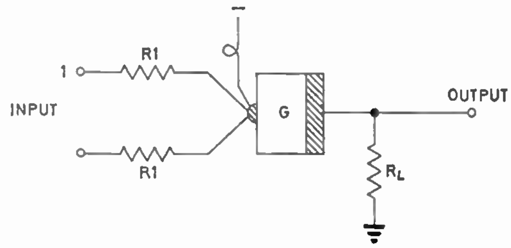


Fig. 9. A Gunn-effect AND circuit.

Table 3
Truth table of an inhibitor

A	B	Output
0	0	0
1	0	0
0	1	1
1	1	0

Table 4
Truth table of AND gate

A	B	Output
0	0	0
1	0	0
0	1	0
1	1	1

$S = 2\times 10^{-6}$ cm^2, has a power dissipation of:

$$P_d \simeq 15 \text{ mW}$$

The comparison with the power dissipation of modern conventional logic devices is favourable. For example a T^2L circuit has a power dissipation of 10 mW per gate.[8]

3. Further Logic Elements

Figure 8 shows an inhibitor circuit, whose truth table is given by Table 3. A resistor R has been put in series with one of the input terminals of a comparator.

If a voltage with sufficient amplitude is applied to input 2 of Fig. 8, a domain will form and the device gives an output pulse. However, if the same potential is applied to terminal 1, due to a voltage drop across R, the maximum field near 1, will not be sufficient to nucleate any domain and there will be no output pulse for the device. On the other hand, if the voltage is applied to both inputs, according to the comparator property, there will be no output. Hence, the device behaves like an inhibitor.

Figure 9 shows an AND circuit. It employs an ordinary Gunn diode with two resistors R_1 connected to the cathode electrode. If a voltage V_a is applied to one of the inputs only, the voltage across the diode is:

$$V_{D1} = V_a\left(\frac{R_O}{R_1+R_O+R_L}\right) \qquad \ldots\ldots(6)$$

where V_a is the voltage applied at the total device. If $V_{D1} < V_t$, there will be no domain formation and no output for the device. If the same voltage is applied to both inputs simultaneously, the diode voltage is:

$$V_{D2} = V_a\left(\frac{R_O}{R_1/2+R_O+R_L}\right) \qquad \ldots\ldots(7)$$

It can be seen, that $V_{D1} < V_{D2}$, and if $V_{D2} > V_t$, the device will have an output. The device operates as an AND gate whose truth table is given in Table 4.

4. Some Examples of Logic Circuitry

The elements described so far should enable one to develop logic circuits.[9] Two examples, namely a simple sequential adder and a shift register, will now be given.

As the pulse frequencies are in the microwave region, the devices have to be developed in microstrip technology with the shortest interconnections possible. The first feasibility study will be performed with blocking capacitors in order to separate the signal currents from the biasing. However, ultimately the circuits will have to be designed without these capacitors in the same way as conventional logic, because capacitors will limit the frequency of operation.

For the new logic circuitry, short delays are required with a time-constant equal to the duration of Gunn-effect pulses. These can easily be achieved by short stretches of microstrip line. For example, 8 mm length with semi-insulating GaAs as dielectric gives 100 ps delay.

As shown in the previous sections, a chain of pulse processing devices has to be arranged in such a way that the voltage pulse, produced by a domain, is of correct polarity for the nucleation of a further domain in a subsequent device. There are many more possibilities than are shown in Fig. 2. However, we are giving the following examples on the basis of this technique.

4.1. *Sequential Adder*

Figure 10 gives the circuit of a sequential adder. The circuit has two input connections which are coupled to the two cathode electrodes of a Gunn-effect comparator Co1. The required input-signal pulses have to have negative polarity. The input terminals are also connected to an AND A1, which is in parallel to Co1. As soon as only one pulse signal is received, a domain is only triggered in Co1. On the other hand, two simultaneous input pulses produce a domain only in the Gunn diode of A1. The output of Co1 is connected via a pulse-reversing diode D1 to one input of a second comparator Co2, and the output of A1 is fed via a delay d1 and another pulse-reversing diode D2 to the other cathode of Co2. The pulse produced in Co1 generates immediately a domain in D1 and in Co2, whereas a pulse from A1 is first stored in d1 until the next digits arrive at the input of the adder. If this subsequent digital signal contains only one input pulse, an output signal originates from Co1, at the same time as the pulse from the previous digit is released from the store d1. As both signals arrive simultaneously at Co2, no domain is produced in the second comparator. However, as seen from Fig. 10, the two connections to D1 and D2 are also joined to a second AND A2 in parallel, where the two simultaneous pulses generate a domain in D4. The output of A2 is stored in another delay d2. Now we assume that the next digital input to the adder is again one single pulse producing a domain in Co1. This gives a signal to D1 at the same time as the previous pulse from the store d2 is released and transmitted via the pulse-reversing diode D5 to D2. Again, two input signals appear simultaneously at Co2, where no domain is produced, whereas a domain is triggered in A2. If the next digital input does not show any pulse signal (i.e. representing two zeros), the pulse stored in d2 will travel via D5 and D2 to the comparator Co2, where an output pulse is finally produced, as no simultaneous pulse arrives from D1. The total digital output from Co2 is now 1000, which represents the sum of the two digital inputs.

The reader may try any other summation. The basic property of this adder is its monostable operation which requires accurate timing for all the devices involved. This coincidence requirement seems at first a formidable problem. As Kurokawa[10] has shown, the growth rate of a domain is given by the dielectric relaxation time multiplied by the ratio of low field mobility to the modulus of the negative, transferred-electron mobility. This growth rate is also responsible for the time required for a signal pulse to nucleate a domain. For samples of 150 μm length, a material resistivity of 10 Ω cm can be employed, so

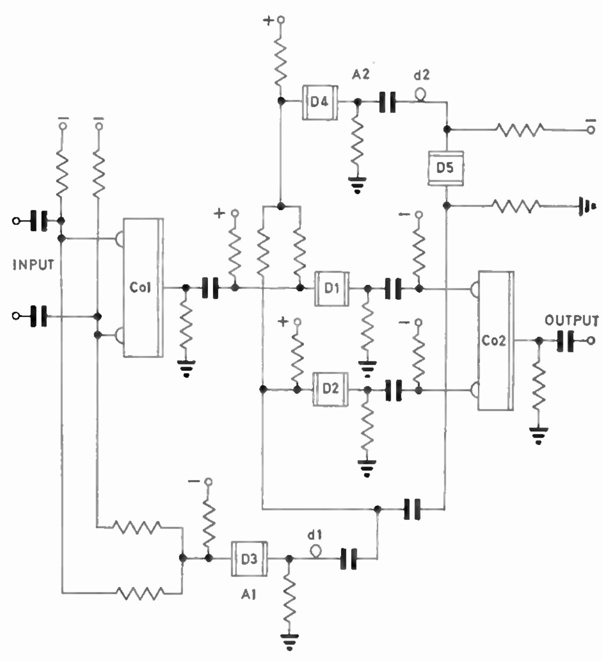


Fig. 10. Binary sequential adder.

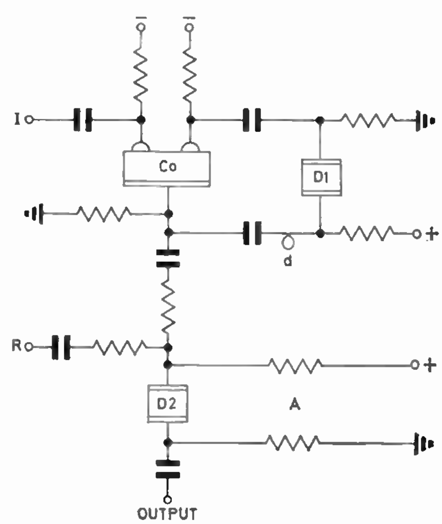


Fig. 11. A memory loop.

that the nucleating time is about 30 ps. This means a pulse coincidence of at least 2%. This theoretical figure represents a difficult task, which should, however, be solvable with microstrip technology.

4.2. *Shift Register*

A further logic circuit is presented finally, which would give a computer element whose increase in speed would represent an essential advantage. Figure 11 shows a loop which can be employed as a memory device. A negative pulse signal can be fed-in via the connector I, and a domain will be nucleated in the comparator Co. This produces a positive output pulse, which is first stored in the delay d, is then reversed in polarity by the diode D1 and is finally applied to the other cathode electrode of Co. This pulse rotates around the loop and can be read-off by a 'read-off' signal applied at R. This will generate a domain in the diode D2 and the AND gate, A, when read-off signal and rotating pulse appear simultaneously at the input terminals of A. The stored information can be read-off at the output terminal. This process does not destroy the stored information which can be read-off many times. The stored pulse can, however, be erased if another pulse is applied at I at the same time as the stored signal arrives from D1 at the other comparator cathode. This prevents domain nucleation in Co.

This memory cell can be employed in the way shown in Fig. 12 to form a shift register. For this purpose, the output of D1 is also connected to a second AND gate, A2, whose other input connection is joined to the input terminal I. The first pulse at I introduces a signal into the first loop, whereas no domain is formed in the diode of the comparator A2. A second digital pulse at I erases the rotating pulse of the first loop, and produces a pulse at one of the inputs of A2 at the same time as the rotating pulse from D1 arrives at the other terminals of A2, before it is erased by Co. A domain is therefore nucleated in A2, which gives a pulse signal via a pulse-reversal diode D4 to the second memory loop. One has now stored the information 10, where the last digit (0) can be 'read-off' from the first loop and the second digit (1) from the second loop. Further input signals at I produce a corresponding adding action in the row of loops of Fig. 12. A third digital pulse introduces a pulse in the first loop only. A fourth input pulse erases the pulses in the first and second loop and introduces a signal into the third loop. This represents now the digital information 100.

This shift register can also be employed for ultra-fast counting and should be of particular interest for applications such as nuclear counters.

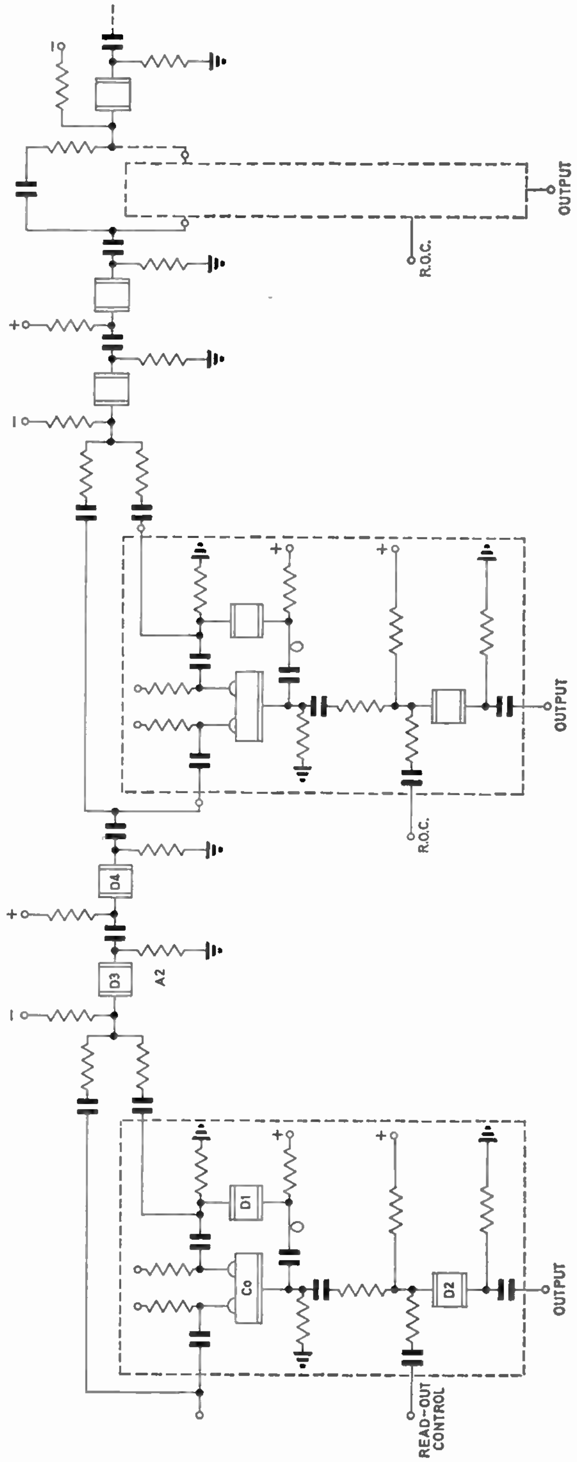


Fig. 12. Circuit of a shift register.

5. Conclusion

It has been shown how the new logic devices are based on the domain control of Gunn-effect elements with resistive loads. They will produce logic circuitry with considerably increased speed of operation and reduced overall size. The time-constants involved are reduced by more than one order of magnitude as compared with the most advanced conventional transistor and diode circuitry and the power dissipation is reasonably low. An essential aspect of these new devices is the possibility of developing simple elements with manifold properties, which can otherwise only be achieved by complex conventional circuitry. This inherent advantage of the Gunn effect is given by the many possibilities of controlling the nucleation, extinction and growth of the travelling domain. A further essential property is the gain mechanism involved in domain dynamics, so that no amplifiers are required between the stages of the logic circuitry.

6. Acknowledgments

The authors are grateful to Dr. S. Mahrous (University of Sheffield) and to Mr. I. Bott (Royal Radar Establishment) for useful discussion and also to the Services Electronics Research Laboratory, Baldock, for the supply of suitable material.

7. References

1. Copeland, J. A., 'Stable space-charge layer in two-valley semiconductors', *J. Appl. Phys.*, **37**, pp. 3602–9, August 1966.
2. Sandbank, C. P., 'Synthesis of complex electronic function by solid-state bulk effects', *Solid State Electronics*, **10**, pp. 369–80, 1967.
3. Shoji, M., 'Functional bulk semiconductor oscillators', *Trans. Inst. Elect. Electronics Engrs on Electron Devices*, **ED-14**, No. 9, pp. 535–46, September 1967.
4. Copeland, J. A., Hayashi, T. and Uenohara, M., 'Logic and memory elements using two-valley semiconductors', *Proc. I.E.E.E.*, **55**, pp. 584–5, May 1967 (Letters).
5. Copeland, J. A., 'Bulk negative-resistance semiconductor devices', *I.E.E.E. Spectrum*, pp. 71–77, May 1967.
6. Hartnagel, H. L., 'Digital logic-circuit applications of Gunn-diodes', *Proc. I.E.E.E.*, **55**, pp. 1236–7, July 1967 (Letters).
7. Izadpanah, S. H. and Hartnagel, H. L., 'Experimental verification of Gunn-effect comparator', *Proc. I.E.E.E.*, **55**, p. 1748, October 1967 (Letters).
8. Carroll, W. N. and Jenny, F. F., 'The airborne 4Pi computer', *Electronics*, **40**, No. 5, pp. 171–6, 6th March 1967.
9. Hartnagel, H. L., 'Some basic logic circuits employing Gunn-effect devices'. *Solid State Electronics*, **11**, pp. 568–72, 1968.
10. Kurokawa, K., 'The dynamics of high-field propagating domains in bulk semiconductors', *Bell Syst. Tech. J.*, **46**, pp. 2235–59, December 1967.

Manuscript received by the Institution on 12th June 1968. (Paper No. 1216/C105.)

Radio Engineering Overseas . . .

The following abstracts are taken from Commonwealth, European and Asian journals received by the Institution's Library. Abstracts of papers published in American journals are not included because they are available in many other publications. Members who wish to consult any of the papers quoted should apply to the Librarian giving full bibliographical details, i.e. title, author, journal and date, of the paper required. All papers are in the language of the country of origin of the journal unless otherwise stated. Translations cannot be supplied.

INTEGRATED CIRCUITS

The May 1968 issue of the *Proceedings of the Institution of Radio and Electronics Engineers Australia*, contains a series of papers on the physical principles and circuit applications of integrated circuits. The following are some of the papers in this issue.

'Hybrid thin film micro-electronics—a quick and flexible method of micro-circuit construction', A. J. Marriage and R. N. Wheaton, pp. 137–42.

Thin film integrated circuits can be made by depositing a pattern of resistive, dielectric and conductive materials on to an insulating substrate. The advantages and limitations of hybrid thin film microcircuits, together with a discussion of materials and techniques are described. A description of a manufacturing process for producing a resistor-conductor network using a nickel-chromium resistor material is given. The limitations on range, tolerance and power dissipation of various components are discussed and the most suitable circuit application areas are indicated and illustrated with some examples.

'Gyrator and state-variable results for linear integrated circuits', R. W. Newcomb and B. D. O. Anderson, pp. 143–50.

In a survey paper various methods are discussed for the design of linear integrated circuits. Basic to the methods are the practical availability of integrated operational amplifiers and gyrators for which the underlying theories of state-variable cascade synthesis methods are applicable.

'An introduction to analogue techniques in m.o.s.t. integrated circuits', D. E. Hooper, pp. 151–63.

Another paper reviews the essential properties of the metal-oxide silicon transistor (m.o.s.t.) which is fabricated with a simple, single-diffusion technology and shows how these properties can be utilized in the realization of some basic analogue circuit functions. In particular, the high degree of linearity achieved in analogue switches and simple non-feedback gain stages is of theoretical interest and practical value. The overall aim of this introductory treatment is to stimulate the interest of circuit designers in the hope that they can contribute new circuit techniques that are compatible with the new disciplines imposed by m.o.s.t. technology.

'Some applications of monolithic circuits in telemetry', J. R. Groves, pp. 164–66.

The use of monolithic circuits has made a considerable difference to the accuracy and possible complexity of the electronics which can be accommodated in a rocket and some applications of these circuits to airborne instrumentation are described in a further paper. Included are thermocouple and strain gauge amplifiers, an analogue-digital converter and a charge amplifier used with piezo-electric transducers, all of which are designed for use in a free flight rocket programme.

'The process steps for integrated circuits and their relation to customer requirements', J. C. van Vessem, pp. 170–76.

A careful analysis of customer requirements for integrated circuits with respect to price, performance and quality leads to a set of basic conditions for the designer and the technologist, which can be related to the essential process steps. These steps are critically examined in another review-paper in relationship to the requirements mentioned above.

'Preparation of indium antimonide films', B. M. Bartlett, pp. 177–82.

Advances in the area of semiconductor thin film technology and surface phenomena have stimulated the development of thin film transistors that are fully compatible with thin film microcircuits. The preparation by evaporation techniques of thin film compound semiconductors on amorphous substrates for this purpose is of considerable technical interest. Most compound materials decompose to varying degrees when heated to obtain evaporation and because there is often a great difference in the volatilities of the constituents, the composition of the condensed film usually differs considerably from that of the compound. A brief review is given of techniques which overcome this problem and permit thin, stoichiometric layers of compound semiconductors, III-V compounds in particular, (i.e. compounds formed by combining an element from Group III of the periodic table with an element of Group V) to be made. A brief summary of results which have been obtained with thin indium antimonide films by various workers in this field is also given. Progress with the preparation of indium antimonide films has been such that virtually single crystal films can now be made.

'An integrated circuit time code generator', D. N. Warren-Smith, pp. 190–200.

A time code generator, which utilizes integrated circuits and etched circuit cards throughout, is described. Design principles are discussed with reference to block diagrams to illustrate the inter-relationship of the different parts of the device.

The technique used for obtaining synchronization between the internally generated time and an external time datum is described.

Details of the logical design of a binary coded decimal to serial decimal code converter and of a comparator circuit are given in an appendix.