Founded 1925
Incorporated
by Royal Charter 1961
"To promote the advancement of radio, electronics and kindred subjects by the exchange of information in these branches of engineering."

VOLUME 41 No. 7

JULY 1971

THE RADIO AND ELECTRONIC ENGINEER

The Journal of the Institution of Electronic and Radio Engineers

Whither Quality and Reliability?

Earlier this year, after effecting a change in the terms of reference and funding of the British Productivity Council, the Government decided to discontinue the Grant-in-Aid to the National Council for Quality and Reliability. The following appraisal of the work of N.C.Q.R. during its ten years and some thoughts on its future role has been contributed by the Institution's representative on N.C.Q.R. since its formation in 1961.

IN the course of one decade N.C.Q.R., in the estimation of many, has fulfilled its primary objective of playing a significant part in reminding British industrial and commercial life of the imperative need for the country's products and services to aspire to the highest levels of quality and reliability. The Council's resolution to terminate N.C.Q.R. on 30th April 1971, at least in its original form, was not brought about by misplaced belief that nothing more remained to be done, but by Government decision made known at short notice that a grant-in-aid from public funds was not justified in the present economic climate. Prior to 1st April 1970 N.C.Q.R. was serviced by the British Productivity Council who had fostered its development from the restricted and specialized field of its own sub-committee on Quality Control. Thus the direct grant-in-aid was very short-lived. In retrospect it is only too easy to challenge the wisdom of the Council in relying on outside resources; in fairness it must be said that the speed with which public funds were both supplied and withdrawn was inevitably an unsettling factor.

At the present moment a group of members of the Executive Committee of N.C.Q.R. are studying the case for re-establishing a self-supporting national body and the ways and means of so doing. The writer is convinced of the need for such a body. But it must restrict itself to the essential task appropriate to a national focal point and clearing house. It is not suggested that past 'missionary' work was inappropriate but not infrequent signs of quality being treated by some as just another specialized technique have been deplored. Two factors remain to be resolved which can have some bearing on the present considerations. The most important of these is the strength of local productivity associations throughout the country after the withdrawal of the British Productivity Council's grant-in-aid. The second factor is more speculative—the outcome of the Mensforth Committee's forthcoming report on the means of authenticating the quality of engineering products and materials.

To return to the heading above these words and firstly to simplify it by omission of the reference to reliability. Quality is best understood as an all-embracing philosophy of which the technical achievement of reliability is a part. In parentheses can be mentioned the writer's suggestion made some time back that a better title for N.C.Q.R. would be the National Quality Council. The simplified form 'Whither Quality?' clearly is a purely rhetorical question which neither demands nor elicits an orthodox answer.

Quality is a way of life—personal, institutional and national. N.C.Q.R's last propaganda poster declaimed 'Quality Rewards' and depicted John Bull grasping the monetary gain. An even more vital reward is that of personal satisfaction. A decade's association with N.C.Q.R. on the Institution's behalf has done little to change a personal, basic view on quality, the practical origins of which can be credited to the leadership of one who, many years later, was to become one of the Institution's highly regarded Presidents—Mr. Leslie McMichael. May one hope that the Institution's inherent interest in the techniques of Q & R will continue to encourage this long standing concern with good engineering?

F. G. DIVER

Contributors to this issue

Mr. B. M. Sosin was born in Cracow, Poland, and served during the war with the Polish Army Signals Corps, after which he studied at London University where he was awarded a first class honours degree in engineering. In 1949 he joined the Marconi Company and was engaged in the design of high-power v.h.f. filters. From 1957 he developed transmitters and associated equipment, in particular the technique used in the 'ghost-less' television transmitter, and wideband distributed power amplifiers. Later he was in charge of a design team working on modulation and drive equipment for communications transmitters and leader of the Advanced Development Group. At present he is Engineering Manager responsible for communications transmitters and new receiver techniques within Marconi Communication Systems Ltd.

Mr. Ray Knight is currently employed by Thames Television Ltd. as a member of the Engineering Research Group. At the start of independent television he joined Associated Rediffusion, later going to ABC Television. His activities have been centred mainly on studio vision-control techniques and he was responsible for setting up a system for testing the now overtaken image-orthicon pick-up tube, and for introducing one-man vision-control into independent television. He is now mainly concerned with the operational problems of colour television.

Dr. P. J. Ellis (Member 1967, Graduate 1962) received his technical education at Redhill Technical College and Brighton College of Technology, while working at Mullard Research Laboratories as an electronics apprentice and later as an engineer. In 1962 he joined the Space and Weapon Research Laboratory of Elliott Automation, working on instruments with vacuum physics and satellite applications. After a year as Chief Development Engineer at the North Hants Engineering Co., he received an appointment as a senior research assistant at the J. J. Thomson Physical Laboratory, University of Reading, joining a group engaged in the design of instruments for the *Nimbus* weather satellites. The group has since moved to Heriot-Watt University, Edinburgh. Dr. Ellis is the co-author of a number of papers on infra-red and satellite instruments. He has recently been awarded the Ph.D. degree of Reading University.

Dr. C. H. Clayson gained his B.Sc. in electronic and electrical engineering at the University of Birmingham in 1963. He stayed at Birmingham for a further four years to do research into laser communications systems, mainly concerned with electro-optic modulators and beam waveguides. A paper on some of the aspects of this work was published in the *Journal* in June 1968. In 1967 he joined the National Institute of Oceanography and a year later he was awarded his Ph.D. degree by the University of Birmingham. Dr. Clayson now leads a section of the Applied Physics Group at the N.I.O. concerned with general oceanographic electronic equipment, ranging from servo controls to conductivity measurements.

Mr. Norman D. Smith started work as a laboratory assistant at the Admiralty Research Laboratory, Teddington, in 1940. From 1942 to 1947 he served in the Royal Air Force working on radar. On demobilization he joined the Wave Research Group at the Admiralty Research Laboratories. Two years later he went to the National Institute of Oceanography on its formation and has worked on the development of oceanographic instruments since that time. Mr. Smith is now a section leader of a group within the Applied Physics Department. He was awarded the M.B.E. in the New Year Honours List in January 1971.

Mr. John F. Roulston was born and educated in Northern Ireland. He graduated with B.Sc.(Hons.) in electrical engineering from the Queen's University of Belfast in July 1970; he was awarded an I.E.E. Prize on the recommendation of the Academic Council of the University. During his university course he gained industrial experience with Ferranti Ltd, Edinburgh, through vacation training and he has now joined that company as a design engineer in the Electronic Systems Department of the Radar Group.

Mr. J. F. H. Binns received his main technical education at the Mid-Essex Technical College. He joined Marconi's Wireless Telegraph Co. in 1953 and he has continued with the Company up to the present time. He is now a Section Leader in the Television Transmitter Department of the Broadcasting Division of Marconi Communication Systems Ltd.

U.D.C. 551.465:621.398

Recent Advances in Wave Buoy Techniques at the National Institute of Oceanography

By

C. H. CLAYSON, B.Sc., Ph.D.†

and

N. D. SMITH, M.B.E.†

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

The well-established *Doughnut* and *Cloverleaf* wave buoys developed by N.I.O. are described briefly and details of a disposable wave buoy, at present under development, are given. Because of certain operational difficulties experienced with the *Doughnut* buoys, a digital telemetry system has been developed and this is described in full detail. The system allows for simultaneous sampling of up to 10 analogue channels, at 0·25 second intervals, with 0·2% accuracy.

1. Introduction

The use of the *Doughnut* and *Cloverleaf* buoys in wave measurement has been described by Cartwright and Smith[1] and will not be repeated here but a brief description of the sensing unit and recording system is of relevance to the following discussion. The sensing unit consists of a vertical-seeking gyro with potentiometer pick-offs on the gimbal axes, an inductive type of accelerometer on top of the gyro motor to measure vertical acceleration and a capacitance type of compass to give the directions of the gimbal axes. It is hoped to replace the capacitance compass by a 2-component flux gate sensor during 1971 since it has not proved entirely satisfactory due to ambiguity over the 'dead zone' of 345°–360°. The capacitance compass forms part of the tuned circuit of a 0·5 MHz oscillator. The outputs from this oscillator and from a stable reference oscillator produce a beat frequency signal, varying between 5 kHz and 45 kHz, which is fed to a diode pump producing a direct voltage varying linearly between 0–2 V for the complete swing of the compass.

The potentiometer pick-offs on the gyro gimbals are supplied with 12 V stabilized d.c. and are used as half of a bridge circuit. The accelerometer requires 10 V r.m.s. at 2 kHz and is energized by a simple oscillator running off the 12 V supply. The accelerometer output is detected and balanced against the output from an identical detector, fed by a potentiometer across the 2 kHz supply, resulting in a d.c. output varying between $\pm 0{\cdot}4$ V over the acceleration range of $1 \pm 1\ g$. In the original form of the equipment, the gyro and electronics are housed in a high tensile aluminium HE 30 WP canister, which is connected to the recording instruments on the ship by an 18-core cable. A polypropylene rope is lashed to the cable at approximately 2 m intervals and expanded polystyrene floats are tied on at the same intervals. In addition to the analogue channels, the cable supplies power, separately, for the electronics unit and for the gyro. On board ship, the analogue signals are fed, after appropriate filtering and scaling, into a commercial data logger which sequentially scans up to 16 channels per second and which drives a 5-hole paper tape punch unit. In addition to this record, an ultra-violet recorder is used to obtain an analogue record, mainly as a check on the quality of the signals and to resolve any ambiguities in the compass record, but also as a safety measure in the event of a data logger failure. It is, however, both laborious and expensive to digitize such a record—approximately 33 000 conversions are necessary per record.

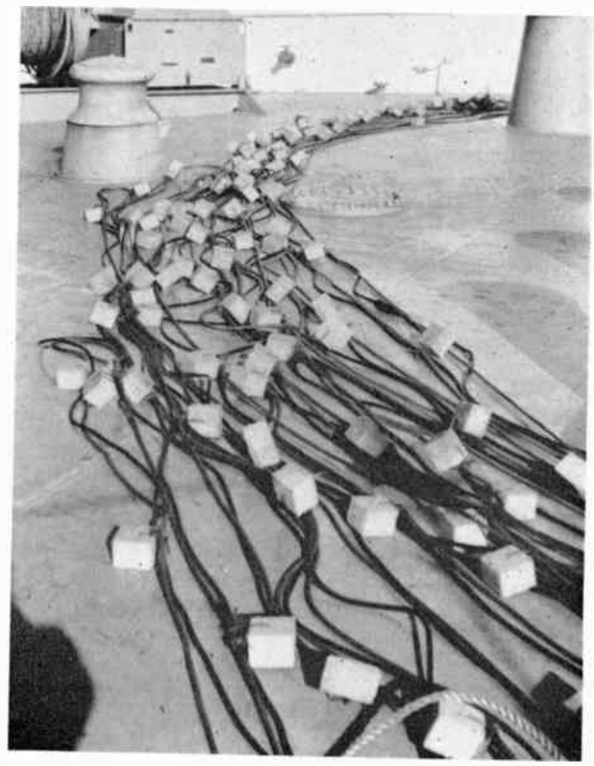


Fig. 1. Conventional cable, supports and floats.

2. Design of the New System

Although the equipment has been used successfully in its original form, described above, for many years, it has some practical disadvantages. The main difficulties are associated with the complex cable. The rigging of the cable, rope and floats takes up a lot of deck space and is a lengthy procedure. Figure 1 shows the cable with floats laid out. The resulting assembly is prone to tangles both in and out of the water; also, the floats are frequently damaged as the cable passes over the ship's side. Since the cable is not armoured, for reasons of weight and flexibility, its life is limited since cuts in the

† National Institute of Oceanography, Wormley, Godalming, Surrey.

outer sheath result in faults due to water creeping along the cable and eventually causing short circuits at the connectors. Not only are these faults troublesome to eliminate but they may also cause undetected changes in the calibration of the analogue channels before the catastrophic failure occurs. The deficiencies of the compass have already been mentioned.

The rigged cable and the shipborne recording equipment are bulky, posing problems when the system is used on small ships. The paper tapes for one cruise may occupy as much as 30 ft^3: this also poses problems of archive storage space. Finally, with regard to the processing of the data, the sequential scanning at 16 channels/second results in additional expensive computation since the time lags between the channels must be eliminated by interpolation techniques. A simultaneous sample/hold unit would save this additional expense.

In November 1969 it was decided to investigate systems of telemetry which would ease some of these difficulties. Radio telemetry is attractive since the problems of the cable would disappear and it would be much simpler to keep the ship stationed correctly with respect to the buoy: at present this requires considerable skill on the part of the ship's officers since the ship must not be in such a position that the waves at the buoy are affected due to diffraction or reflexion, nor must the cable pull on the buoy excessively. However, there are problems of power storage and recovery in the case of a radio buoy, and though these should be soluble, it was decided to telemeter over a cable initially but using a system which could be readily adapted to radio transmission later.

The approach adopted was to develop a multiplexer that would allow the use of a much simpler cable, such as the 4-core neutrally buoyant cable used by the N.I.O. for towed hydrophones. This cable has cores of fairly high resistance, 56 Ω for a 300 m length, so that power transmission to the buoy must take place at a minimum of 100 V in the interests of economy and regulation. In practice, it is convenient to use 240 V a.c. as this eases the problems of power conversion. Initially it seemed that frequency division multiplex would provide a simple means of data transmission but this was abandoned on account of the filtering problems and of the total bandwidth which would be required. If, for example, 0·2% accuracy were required with sampling at $\frac{1}{4}$ second intervals, the frequency excursion of each subcarrier must be ± 2 kHz and up to 10 channels would be required, 4 being required for *Doughnut* buoy work and 10 being required for *Cloverleaf* buoy work. These figures for accuracy, sampling rate and number of channels were chosen as being adequate for future wave buoy work but, as the system evolved, it became apparent that it might prove useful in other applications requiring data logging.

Whilst frequency division multiplex is not such an easy solution as might have been thought, many transducers do produce a frequency analogue output (e.g. capacitance compass, f.m. pressure transducer, disposable wave buoy accelerometer) and, because of this, a 'brute force' approach to multiplexing was finally adopted. This approach, whilst in some ways inelegant, has the advantages of simplicity and of computer compatibility. It employs digital transmission of b.c.d. data in serial form. The data signals are generated as follows. Simple, but accurate, voltage controlled oscillators (v.c.o.s) are used to convert all voltage analogue transducer outputs to frequency analogue signals. These signals, together with any directly obtained and suitably scaled frequency analogue signals, are gated into individual 3 decade counter chains for a period of 150 ms in each $\frac{1}{4}$ second. The 12 bit b.c.d. outputs from the counter chains, together with any directly obtained 12 bit parallel b.c.d. transducer output signals, are then serialized sequentially using three dual, parallel 4-bit input shift registers for each pair of counter chains: this serialization occupies the remaining 100 ms in each $\frac{1}{4}$ second. All the necessary gating, preset, reset and clock pulses are generated by a crystal-controlled time-base. From the point of view of servicing, each channel virtually has a complete but very inexpensive a.d.c. of its own and only the time-base is common. Thus, an a.d.c. failure need not be catastrophic, since spare channels may be available and replacements are, in any case, inexpensive. This factor and the flexibility of the system are thought to be worth-while features in comparison with commercially available (and more expensive) modular scanners used with a single a.d.c. The price paid is in power consumption, although this is not of overriding importance in a tethered buoy system. For the 10-channel a/d multiplexer built, the power requirements are 160 mA at +12 V, 200 mA at −12 V and 1·8 A at +5 V, using standard 74 TTL. For 4 channels, as in the case of the *Doughnut* buoy, the current requirements drop to 64 mA, 80 mA and 0·9 A respectively. These figures must also be viewed in relation to the 14 W consumption of the gyro unit. A small power saving could be gained by the use of low-power TTL but counters and shift registers were not then available in that form. If very low consumption was necessary, complementary symmetry m.o.s. logic could be used but some re-design would be necessary and the increase in component costs would be very considerable.

The design of the v.c.o.s, discussed in detail below, will probably be improved upon when time permits. The centre frequency of each oscillator is nominally 3·993 kHz and the full-scale frequency shift is $\pm 2{\cdot}667$ kHz so that the count accumulated over the 150 ms period is 599 ± 400 full scale. The minimum v.c.o. frequency was limited by feedback problems so that the maximum possible count range of 500 ± 500 could not be achieved and $\pm \frac{1}{2}$ l.s.b. corresponds to 0·125% of full scale instead of 0·1%. The reduction in resolution is not significant but it is hoped to rectify this state of affairs in future. It might be imagined that the averaging of the analogue signals which results from the 150 ms sampling period will cause errors in the wave spectra. However, it is easily shown that, for an analogue signal of frequency ω radians/second, the effect of averaging over time δt is a reduction of amplitude by the factor $\sin(\omega\delta t/2)/(\omega\delta t/2)$. For 0·1% error, $\omega = 1{\cdot}03$ rad/s and for 1% error, $\omega = 3{\cdot}26$ rad/s. No phase errors result and the amplitude errors are negligible for most wave measurements.

2.1 *Data Format and Timing*

The format and timing of the transmitted data were chosen so that it is a simple matter to interface the data receiver with the IBM 1800 computers used on the R.R.S. *Discovery* and at the N.I.O. This allows on-line data logging on disk stores when the system is used on the R.R.S. *Discovery*. When the system is used on other ships, the data are recorded serially in NRZ–1 form on an inexpensive two track tape recorder and are replayed via the same data receiver and interfacing unit into the N.I.O.'s computer upon return. The volume of the magnetic tape records can be as little as 2 ft^3 (0·056 m^3) for 300 wave records (cf. 30 ft^3 or 0·85 m^3 for punched paper tape).

The transmission of the ten 12-bit b.c.d. words takes place during the 100 ms period after counting of the v.c.o outputs has taken place, in each $\frac{1}{4}$ s scan. The clock rate is 2 kHz, so that each word occupies 5·75 ms, and 4·25 ms gaps are interposed between words. There is also a 2·25 ms gap between the end of the counting period and the beginning of the 1st word, and a 2 ms gap between the end of the 10th word and the beginning of the next counting period. The interword gaps are desirable on account of the relatively slow response of the computer's digital input circuits and also allow the re-establishment of bit synchronization of the receiver between channels if loss of sync. should occur due to transmission faults. Similarly, the 154·25 ms interscan gaps allow re-establishment of channel synchronization.

It would, of course, be virtually impossible to process the 10 × 12 bits/scan of data reliably without some form of synchronizing or clock pulse train, unless '0's were distinguishable from gaps. This pulse train could take the form $(G1/2+G3/4+G5/6+G7/8+G9/10)$—see transmitter timing chart (Fig. 2). This form of synchronizing waveform suffers from the disadvantage that any difference in time delays in the data and sync. transmission paths, including the magnetic tape recorder record/replay skew errors, will necessitate correction. The circuit shown in Fig. 3 was designed to carry out this function and will tolerate at least ± 200 µs of combined skew and timing jitter. If, however, the synchronizing pulse train, S, takes the form of the inverse code to the data, G, such that if G is, for example, 100100110110, then S is 011011001001, this type of skew correcting network is unnecessary: the maximum tolerable relative delay remains about the same, however. The receiver clock is then generated from $(G+S)$, giving 111111111111, and the noise common to both the G and the S transmission paths can be detected from $(G.S)$. Drop-outs in $(G+S)$ can be detected, within limits, by the error detection/resynchronization circuits described below. Note that both G and S can be transmitted using opposite polarity pulses over a single transmission line and recorded on a single tape track but that a return-to-zero (RZ) mode is then required with possible reduction in reliability.

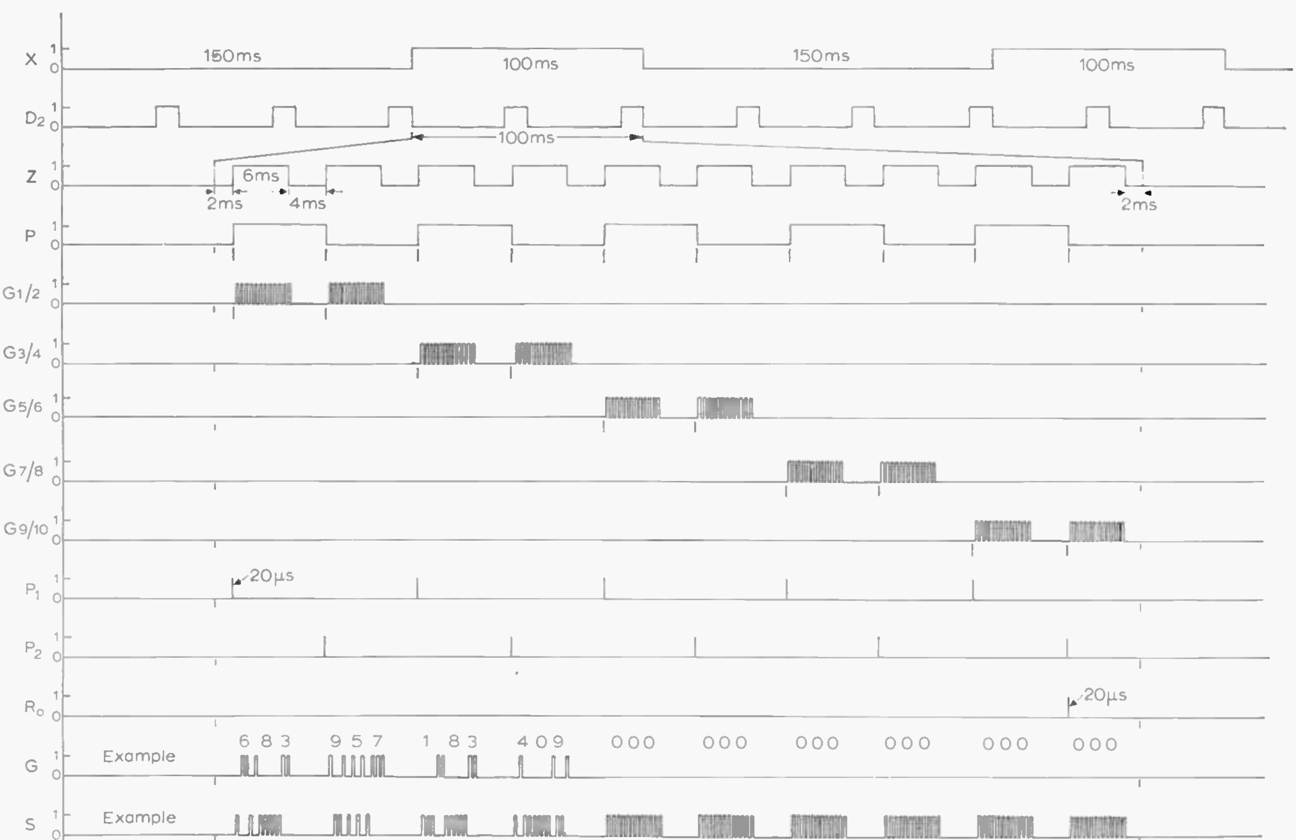


Fig. 2. Transmitter timing diagram.

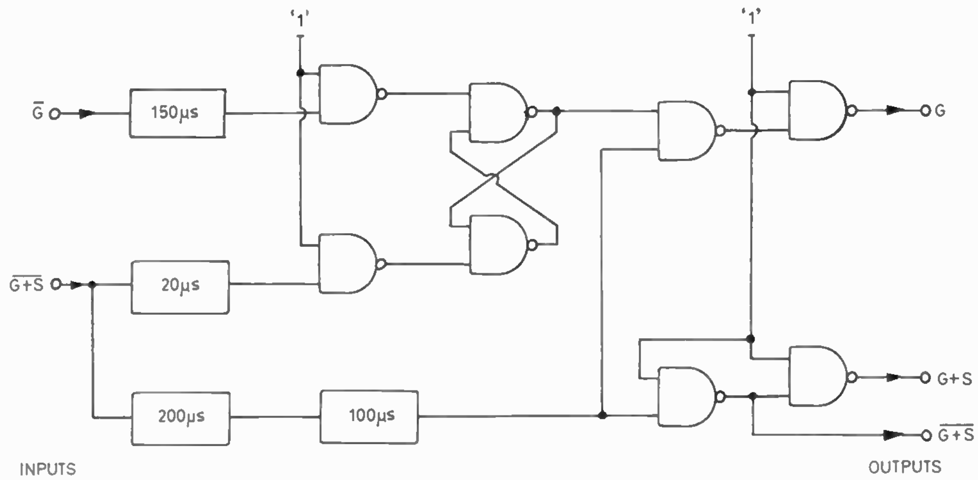


Fig. 3. Gating for skew/jitter compensation.

2.2 *Receiver Facilities Considered Desirable*

The primary requirement is that the receiver must reliably lock on to the transmitted data and sort these into the individual channels. Secondly, it must convert the data into a form suitable for on-line connexion to the computer for storage on disk and all necessary process interrupts and synchronization pulses required by the computer must be generated. Thirdly, it must convert the data into a form suitable for recording on a two-track, continuous, tape deck and be capable of retrieving the data reliably from the tapes. Fourth, it must be capable of selecting and displaying on a numerical indicator display the count, ranging between 199 and 999, representing each channel. Finally, it is desirable that a selected channel be converted back into analogue form for display on, for example, a u.v. recorder. This allows a qualitative check on the operation of each channel. Ideally, one would like to be able to display all channels in analogue form simultaneously, but this is probably not economically justifiable. These facilities are all provided in the system, which will now be described in more detail.

3. Transmitter

3.1 *Voltage-controlled Oscillators* (*v.c.o.s*)

Each oscillator board comprises an input buffer amplifier, a feedback amplifier whose output drives the base circuit of a unijunction transistor (UJT) oscillator, a monostable producing 30 μs TTL compatible pulses, and a diode transistor pump discriminator circuit. The UJT oscillator, by itself, has a highly non-linear frequency/input voltage characteristic, but the use of feedback derived from a linear discriminator, giving an open loop gain of between 1600 and 250, results in adequate overall linearity. The gain of the input buffer amplifier is adjusted so that its output varies between ± 8 V over the full range of the transducer output voltage. The output of the discriminator circuit varies over a similar range for $\pm$ full scale frequency shift, and is approximately 0 V for a frequency of 4·4 kHz. The difference between the buffer amplifier and discriminator outputs is amplified by the feedback amplifier and the resulting output drives the emitter circuit of the UJT oscillator. UJT emitter is connected to a monostable which is

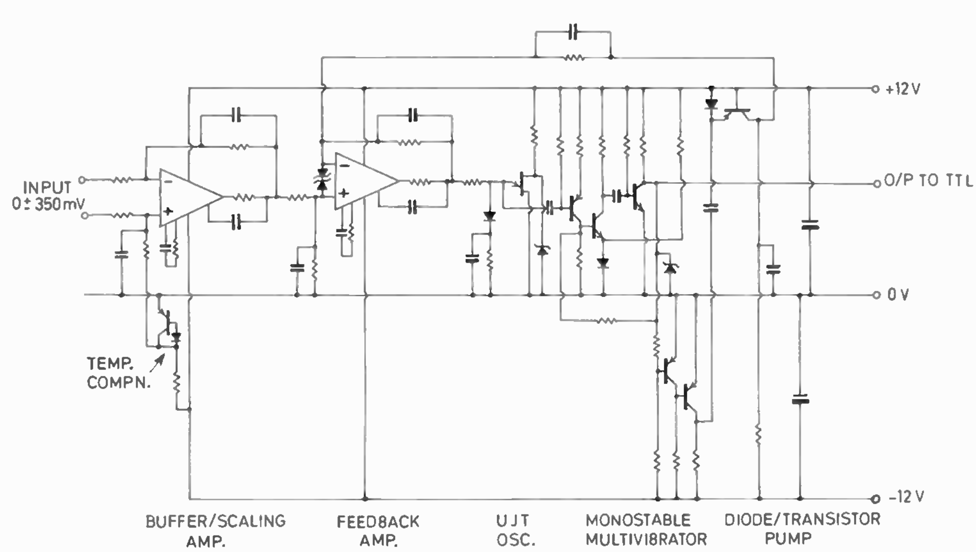


Fig. 4. Voltage controlled oscillator.

triggered each time the UJT fires. When the UJT fires, the lower end of the pump capacitor is switched from the −12 V line up to the 0 V line for 30 μs. It can easily be shown that minimum sensitivity of the oscillator frequency to supply voltage changes then occurs when the discriminator output voltage is $-V_t$ where V_t is the sum of the forward-biased diode and transistor emitter-base junction voltages, and that these junctions are the principal contributors to the temperature coefficient of the oscillator frequency. These facts suggested the simple temperature compensation circuit shown which effectively sums a voltage approximately equal to $-V_t$ with the other voltages at the input to the feedback amplifier (see Fig. 4).

The resulting temperature coefficients of the centre frequency for four v.c.o.s tested over the range 10°C–30°C were respectively 25, 25, 38 and 75 parts in 10^6 per deg C and the supply voltage sensitivities were 20, 25, 50 and 55 parts in 10^6 per mV change in negative supply voltage. The effect of changes in the positive supply voltage was negligible. The effects of negative supply voltage changes on the ratio (frequency shift/input voltage) were all very close to the theoretical value of 90 parts in 10^6 per mV. (These results show that the performance of the v.c.o.s is adequate for this application if the temperature is stable to ±15°C and the negative supply voltage is stable to ±10 mV. The linearities were all better than ±0·1% of full scale frequency shift.)

3.2 *Counters and Serializers*

The monostable outputs from the v.c.o.s are gated, for a period of 150 ms in each 250 ms, into the decade counter chains as shown in Fig. 5. At the end of this period, the accumulated counts will range between 199 and 999 (± full scale) and these counts in 12-bit b.c.d. form are transferred into the dual input shift registers, by appropriately timed preset pulses, and are then serialized by the application of 10 groups of 12 clock pulses at 2 kHz. The shift register outputs are strobed with the gating pulse trains G1/2, G3/4, . . ., etc., and combined into a common signal line, G. In order to simplify the time-base circuitry, the same pair of preset lines P_1 and P_2 is used to preset each shift register chain. If the registers were not cleared before each P_1 pulse, the odd channels numbered 3, 5, 7, 9 would have the data of the even channels 4, 6, 8, 10 (respectively) superimposed upon them, resulting in incorrect data output in these channels. To prevent this from occurring, all of the shift registers are clocked with the same pulse train of 10 × 12 pulses so that data put into the registers by inappropriate presets are shifted out into inhibited gates, thereby effectively clearing the registers. The integrated logic circuits are all standard 74 TTL.

3.3 *Time-base*

The sampling rate must be accurate for wave buoy work and a 4 kHz crystal oscillator is used as a time standard, thereby eliminating the need for any fine trimmers and setting-up procedure. A total of 4 decade counters, 1 flip-flop, 5 monostables, 1 5-bit shift register connected as a ring counter and 44 gates are used to provide all the necessary preset, reset, clock and gating pulses. A number of discrete components are also used to cater for the high fan-outs of the counter reset line R_0 (3 × no. of channels) and of the preset lines P_1 and P_2 (6 × no. of channels, each) (Fig. 6). The timing diagrams (Fig. 2) show how the transmitted pulse format is achieved. Two 20 μs monostables, triggered by a manually switched bistable, are used to initiate the correct operational sequence of the time-base or to reset it in the event of loss of sequence for any reason. A future modification might be to have automatic resetting after the application of the 5 V power so that the time-base can be reset whilst the wave buoy is remote from the ship by temporarily interrupting the power supply.

This would not be possible in a radio or sound telemetering version, which would require a command receiver to achieve this process. However, the loss of correct time-base sequence would only be likely in the case of a very severe mechanical shock which would only be likely whilst the buoy was being handled. The 4-channel transmitter circuitry is split up into 10 printed circuit boards, with 24-way 0·1 in (2·54 mm) spacing, edge connectors, plus the plug-in crystal oscillator package, and occupies a volume of approximately 160 in³ (2600 cm³), including power supplies.

3.4 *Line Drivers and Power Supplies*

The characteristic impedance of the cable used for data transmission (Fig. 7) is highly reactive at the clock frequency of 2 kHz and is frequency dependent in magnitude and, to a lesser extent, in argument. It is not practical to match the cable properly to the receiver at all the frequencies contained within the data spectrum, but an approximate match may be made at 2 kHz and, since the two-way propagation time along the cable is not more than 3 μs, reflexions are not troublesome.

The power is supplied at 240 V, 50 Hz along two cores of the cable. Miniature encapsulated power supply units are used to provide 5 V for the logic circuitry, ±12 V for the v.c.o.s and for the wave buoy circuits and 28 V for the gyro. A two-section LC filter is used to reduce the possibility of impulsive interference if the system is used on a noisy ship's power supply.

4. Receiver

4.1 *Line Receivers*

The lines used for transmission of the G and S signals are terminated for a nominally correct match at 2 kHz and line receivers with a set threshold of approximately half the peak received voltage are used to discriminate against noise picked up during transmission.

4.2 *Time-base Generation and Error Detection/Correction*

The clock for the receiver is simply derived from (G+S) which is fed into a 7492 divide-by-twelve counter. The ABCD outputs from this counter are gated to give (A+B+C+D) as shown in the receiver timing diagram (Fig. 8). Bit errors are detected by gating(A+B+C+D) with the output of a Schmitt trigger, T_1, which goes from

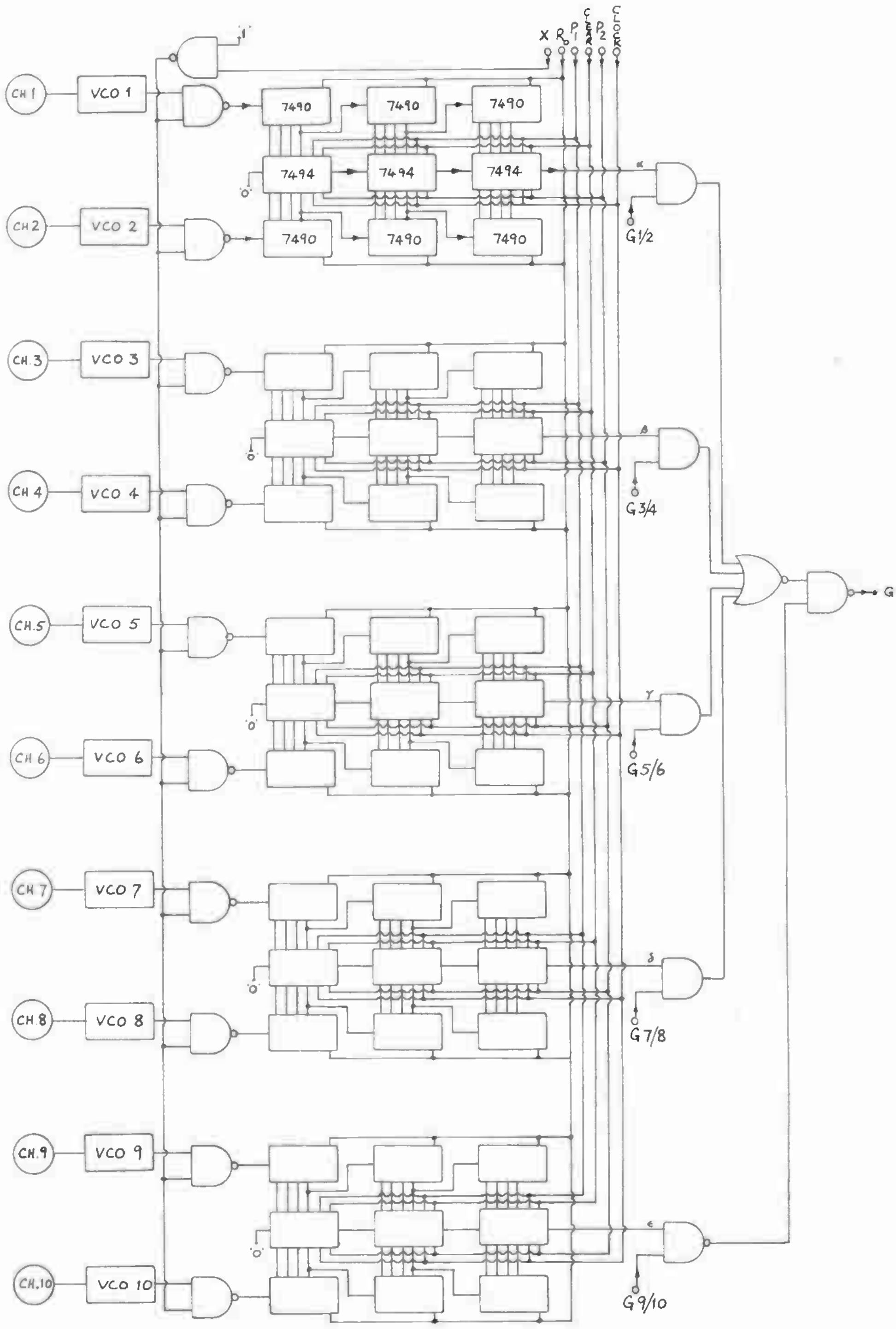


Fig. 5. Transmitter gating, counting and serialization.

'0' to '1' when the output from the bit detector drops to mid-way between '0' and '1' levels (Fig. 9). The rise and fall times of the detector are such that, if the receiver is in synchronism with the transmitter, $(A+B+C+D).T_1$ is always '0'. If there is lack of synchronism, $(A+B+C+D).T_1$ goes from '0' to '1' approximately 1 ms after the end of a group of twelve $(G+S)$ pulses and results in the resetting of the 7492 and the recording of a bit error in the computer disk store together with data bits received up to that time, enabling post mortem fault diagnosis and correction. This is one of the reasons for having 4 ms gaps between the channels. Temporary lack of bit sync. will occur when the equipment is first switched on, unless the transmitter and receiver are reset simultaneously; it will also occur if there are any pulses lost or gained in transmission. In the event of severe loss of data or excessive noise this system cannot work wonders but it will at least ensure a return to bit sync. if the noise abates. A study of what happens under various fault conditions shows that, although bit sync. will be regained, channel sync. may be lost in some circumstances: also it is necessary to achieve channel

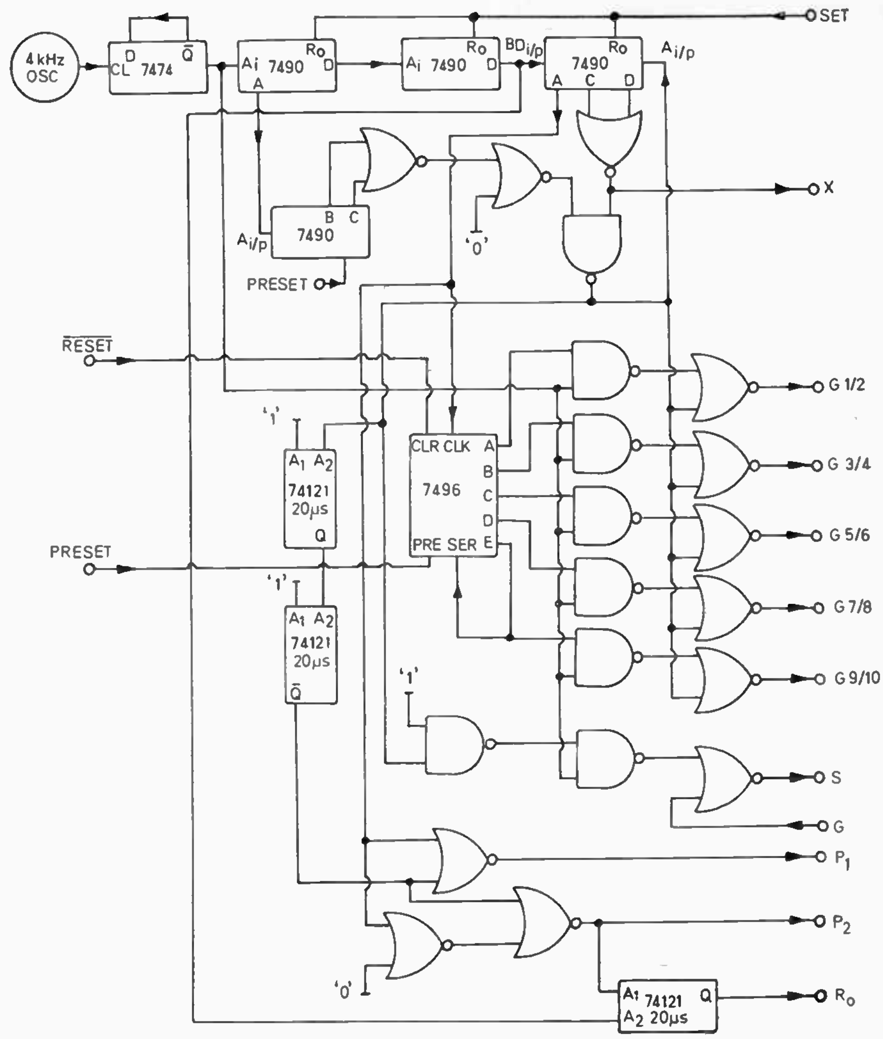


Fig. 6. Transmitter time-base.

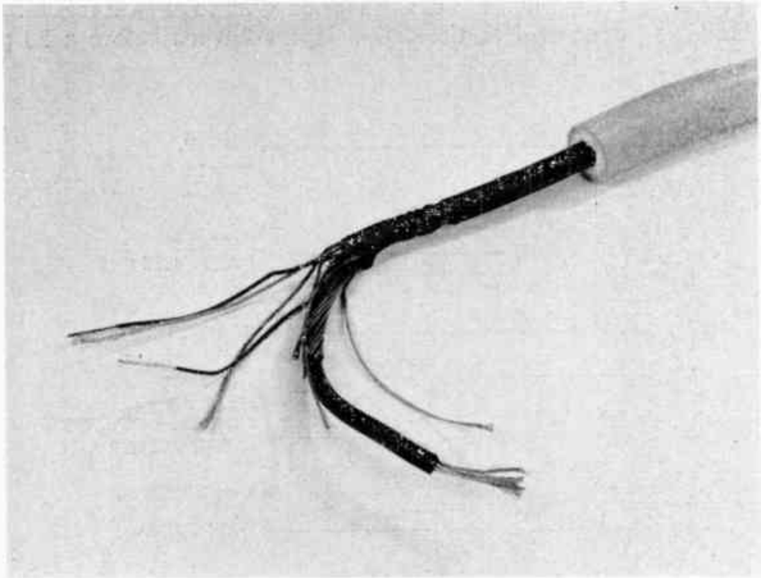


Fig. 7. Buoyant six-core cable.

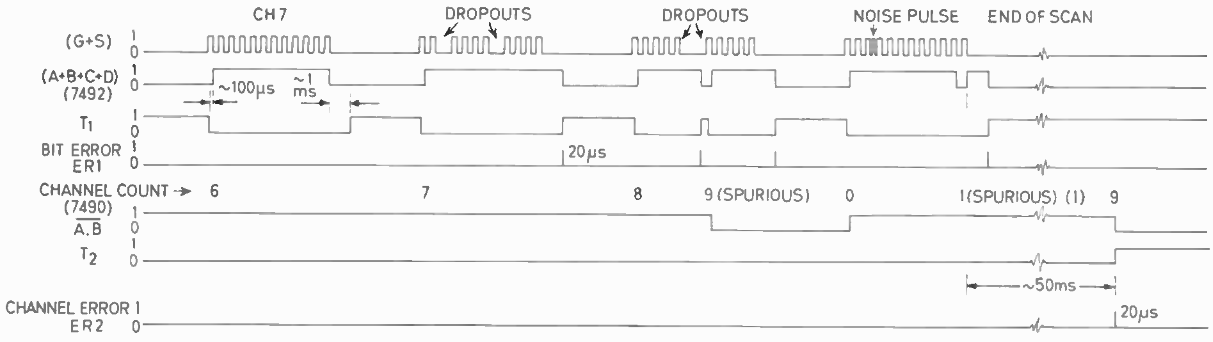


Fig. 8. Receiver error detection timing diagram.

sync. initially after first switching on. Channel sync. correction is achieved by the use of a similar circuit, operating with a decade counter (there being 10 channels), and a detector with a much longer time constant so that its related Schmitt trigger output, T_2, only goes high about 100 ms after the end of the data scan (during the following counting period of 150 ms). This is also shown in Fig. 8. This time a channel error bit is recorded on the computer, at the same time as the first channel of the following scan, so that error location and correction are again facilitated.

4.3 *Channel Selection and Display*

A ten-stage ring counter using two 7496s is used to allow channel selection. The output from the appropriate stage is selected by a 10-way switch and is used to gate the clock, delayed 20 μs after the positive going edges of (G+S). The gated clock pulses so derived are used to shift the serial data G into a 12-bit parallel output register and the b.c.d. outputs from the register drive indicator tubes via 7441A decoder/driver units. The ring counter is reset and the register is cleared when the manual reset is operated and when the channel sync. correction circuit operates. It should be noted that, since there are no buffer stores between the 12-bit register and the 7441As, incorrect codes are applied to the latter for a period of 6 ms in every 250 ms. This causes a momentary flicker in the display which is of no importance. However, the d/a convertor for obtaining an analogue record of a selected channel runs off this same register and, consequently, 'glitches' occur in the analogue output due to the momentarily incorrect input codes. These 'glitches' can be used as accurate timing marks but, if it is required to eliminate them, heavy filtering or the use of buffer stores becomes necessary. Such a modification may be considered desirable.

4.4 *Computer Interfacing*

The system is interfaced so as to allow direct connexion to the N.I.O.'s IBM 1800 ship and laboratory computers. It could probably be adapted to suit other computers having digital inputs for data logging applications, but a brief summary of the interfacing for the 1800 follows as an example of what may be required.

One of the most relevant features of the 1800 digital input circuitry is the filters are incorporated which result in a sensing delay of 2·5 ms for a change '0' to '1' and 5 ms for a change '1' to '0'. However, since there are 4 ms gaps between each word of data it is possible to use the system without any buffer storage in the interface other than a serial/parallel conversion register. The sequence of events is indicated in the timing chart (Fig. 10). Since the register outputs are gated with the channel clock waveform, the digital inputs are set to '0' for a period of 5·5 ms whilst the register is filled: then, when the channel clock goes to '1' the register is connected to the digital inputs 2·8 ms before a 50 μs

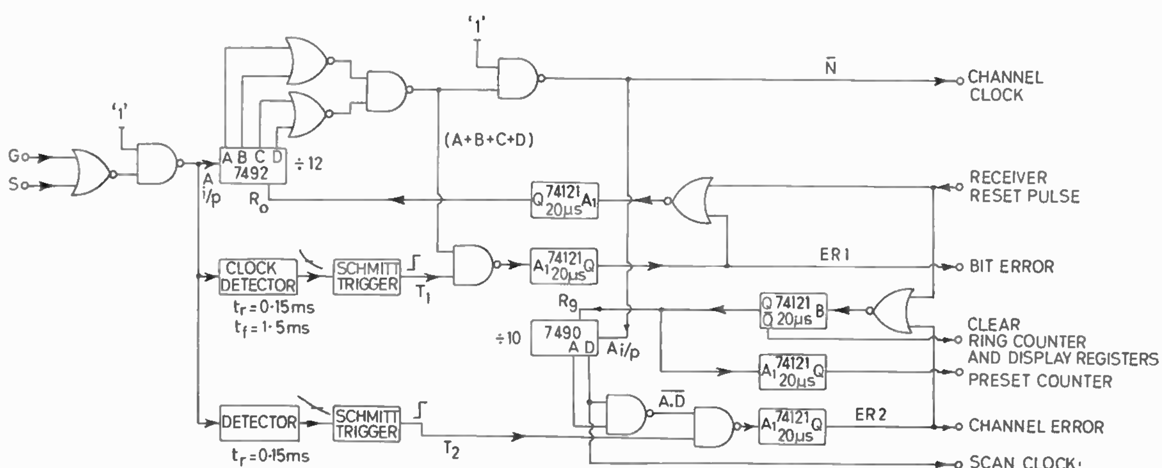


Fig. 9. Receiver error detection and auto-resynchronization circuit.

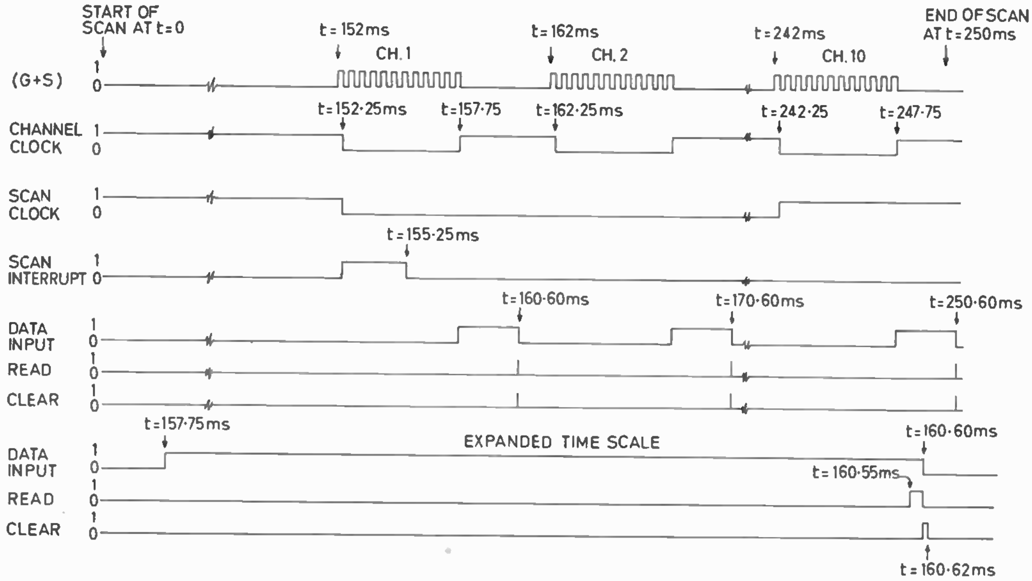


Fig. 10. Timing of data input to IBM 1800 computer.

'read' pulse is sent to the computer to tell it to record the inputs. Therefore the digital inputs will have been either at '0', for 5·5 + 2·8 ms before 'read', or at '1' for 2·8 ms before 'read'. These periods cater for the sensing delays adequately. The register is cleared immediately after the 50 μs 'read' pulse. In addition to the 12 bits of data G, it is necessary to record any bit errors and channel errors and D-type flip-flops are used to store these temporarily until the end of the 'read' pulse. When noise pulses are present in (G+S), the data may continue to be shifted into the serial/parallel conversion register after the channel clock goes from '0' to '1' initiating the sequence of input to the computer. If only one noise pulse occurs, the data in the register will be static 0·255 ms after the channel clock goes from '0' to '1' and the overflow bit is recorded as part of the 'error word' containing the bit error. If more than one noise pulse occurs, the data in the register will not be static for the duration of the sensing delay and incorrect transfer may result. A maximum of three overflow bits is catered for. Also recorded in this word is the output from a D-type flip-flop which is set when (G.S) goes high, this being an error condition.

In addition to the data and to the 'read' pulses, various process interrupts are required: these comprise an interrupt at the start of the record (synchronized to the start of a scan), interrupts (on another line) 5·5 ms before the sequence of input commences in each scan and, finally, an interrupt at the end of the record (synchronized to the end of a scan). The computer can then operate in a time sharing mode but is dedicated to receiving 10 × 2 words of data after each scan interrupt. In some cases of drop-out, when the bit error circuit operates more than 2·8 ms before the end of a 12 bit word is received, more than ten read instructions will result during the scan so that an incomplete scan record will result: a future modification might be to de-restrict the number of words logged by the computer, but to generate an end of scan interrupt. However this fault only occurs if a group of 2 or more of the *first 6 bits* of a (G+S) 12 bit 'word' are absent; if a group of 2 or more of the *last 7 bits* of a (G+S) word are absent, then only those last 7 bits are not recorded. Without going into further detail, it should be apparent that it is possible to make a diagnostic chart for error tracing based on the accurately known timing: the validity of this may be reduced if excessive time jitter is introduced by the magnetic tape recorder, of course.

4.5 *Tape Recorder Interfacing*

A TRD 1/S recorder, manufactured by Tape Recorder Developments Ltd., was used: this incorporates solenoid control of start/stop so that it was possible to synchronize the recorder to start or stop at the beginning of a scan. In practice, this is not essential since the receiver will automatically synchronize itself to the recording on playback after which the 'start of record' interrupt can be initiated manually. The number of scan interrupts applied to the computer can be counted by the computer or on an external counter so that the length of record logged by the computer can be controlled accurately. This is necessary, for example, with some fast Fourier transform programs when the number of terms must be a power of 2. For the initial trials with the equipment 4/4 track tape heads were used and G, S, (G+S) and a start of record pulse were recorded on the four tracks in NRZ-1 form. The only interfacing circuitry required to effect this was two dual D-type flip-flops, arranged to divide by two, and four head drivers switching the head currents between ±5 mA, to ensure saturation of the tape with a reasonable safety margin for loss of contact between tape and head. The same head is used for recording and replay, but a second head is incorporated for read-after-write for qualitative checks on the system operation. At least 25 min change of alignment between record and replay is permissible before skew causes errors, for a tape speed of $7\frac{1}{2}$ in/s (39 cm/s). The raw head output waveform on replay was found to be very satisfactory at the recording speed of $7\frac{1}{2}$ in/s and satisfactory at $3\frac{3}{4}$ in/s. At the latter speed more jitter results for two reasons: Firstly, because of the increased jitter due to the recorder mechanism and secondly, because the output waveform does not pass through the set threshold level as rapidly. The replay circuitry for each track consists of an amplifier with a voltage gain of 1000, a full-wave rectifier and a comparator with the threshold voltage applied to one of its inputs and its output connected to a 100 μs monostable. It was originally thought that the recording of G, S and (G+S) on separate tracks would be useful for parity checking purposes but in view of the extra complication that this entails because of skew and because of the promising quality of the recordings (on Scotch 498 $\frac{1}{4}$-in tape) this may not be necessary and cheaper 2-track heads may be used in future equipment. These should also result in a higher signal/noise ratio.

5. Results

The digital system was used whilst still in the development stage on R.R.S. *Discovery* in April 1970, and on M.V. *Surveyor* in August 1970. A total of 7 hours of recordings were made, using the *Doughnut* buoy, of waves and of ship motion. The data from these recordings have not been fully analysed at the time of writing but preliminary results indicate comparable reliability to the commercial data logger used previously. Approximately 1 million bits are transmitted in the course of a 8192 sample wave record, so that occasional errors must be expected in such a simple system. Improvements, including some simplifications to the logic circuitry and layout, have been made since the initial trials and the system is by no means fully developed. However, the results obtained are sufficiently promising to warrant continuing development. One of the most significant features of the recording system, from the operator's point of view, is the extreme simplicity of the controls: a significant advantage for use at sea.

6. Disposable Wave Buoy

Some progress has been made with the development of a cheap radio telemetering wave buoy, using a novel form of accelerometer. The accelerometer is shown in Fig. 11 in its damping bowl and uses a diaphragm spring as the common electrode in a six-element RC ladder phase shifting network. The six separate electrodes are made in printed circuit form, using a double-sided glass-fibre board with plated-through holes for the connexions to the electrodes. The six 1 MΩ resistors are soldered directly to pads on the back of the electrode board to minimize stray inductance. An RC phase-shift oscillator has been described by Spiess,[2] using a cantilever type of spring and 3 electrodes.

The advantages of the diaphragm spring are considered to be as follows:

(a) It is simple to make a stiff spring with a reliable small electrode spacing, giving high resonant frequency and appreciable air damping at this resonance (60 Hz for the accelerometer built).

(b) The spacing and alignment of the electrodes relative to the diaphragm is set by clamping ring-shaped shims of appropriate thickness between the diaphragm and printed electrode board. Disturbance of the alignment due to handling is not as likely as in the case of a cantilever spring.

However, slight mechanical hysteresis effects occur due to the imperfect constraints at the boundaries of the diaphragm. These amounted to an equivalent zero change of approximately 0·006 *g* after an impulsive loading of 0·5 *g*.

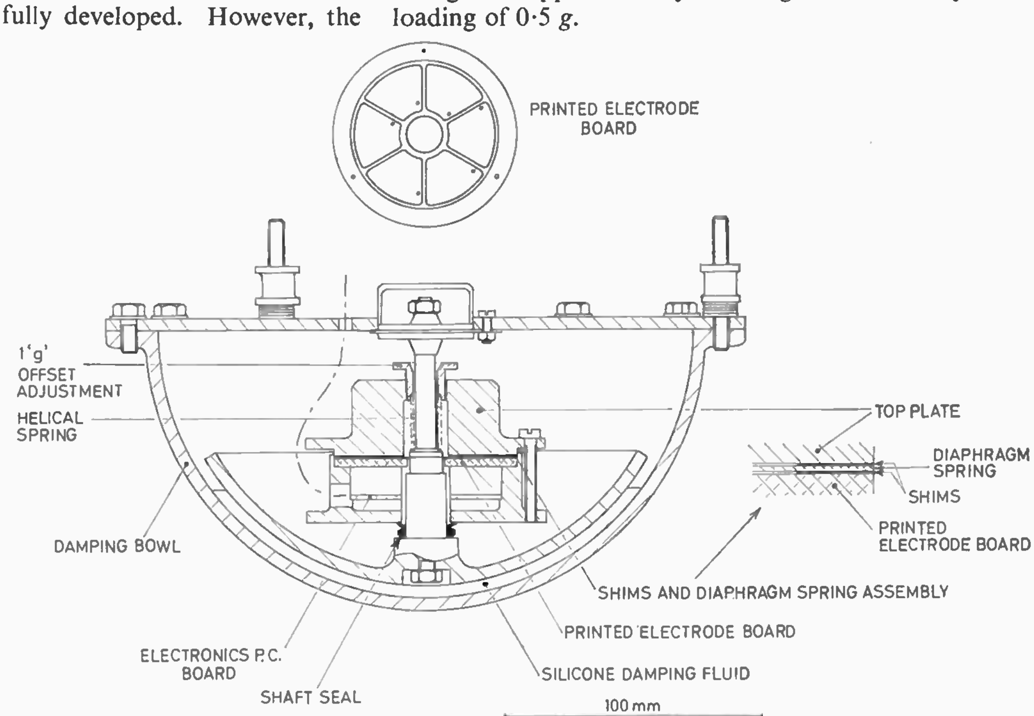


Fig. 11. Diaphragm spring oscillator used in disposable wave buoy.

The advantage of using a six-element ladder, instead of the usual minimum of three elements, is that variations in input capacitance of the maintaining amplifier have less effect upon the frequency of oscillation and that a lower frequency results for the same electrode area and spacing, if the resistors in the ladder have the same individual values. A centre frequency of approximately 3 kHz was required, for reasons of transmitter bandwidth and this was achieved using 0·127 mm beryllium-copper shims of 60·3 mm internal diameter. The sensitivity, using a 0·508 mm thick beryllium-copper diaphragm spring was 400 Hz/g.

A relatively weak helical spring is used to offset the deformation of the diaphragm under normal gravity, as shown in Fig. 11. The compression of the helical spring is adjusted by a screw adjustment, allowing setting to better than 0·005 g. It should be noted that the loading of the diaphragm spring is due to the combined mass of the accelerometer body (in which the maintaining amplifier is encapsulated in epoxy resin), adjusting screw and helical spring.

The accelerometer is suspended from a flexible rubber mounting in a fluid damping bowl so that it takes up the apparent vertical. The errors in the wave spectrum resulting from the use of pendulous accelerometers have been discussed by Tucker.[3] The complete accelerometer unit is encapsulated in an expanded polyurethane float, of 0·915 m diameter, together with the low power radio transmitter and batteries. A loaded whip aerial is used for transmission of the amplitude modulated 27 MHz carrier to the ship. A primary requirement is that the buoy should be capable of being launched from a ship under way. A trial with a dummy float showed that this can be achieved by swinging the buoy on the end of a line through an appropriate angle in a fore and aft plane and releasing it at the bottom of its swing when its horizontal velocity aft is equal to the ship's forward velocity. The buoy then drops with, ideally, zero horizontal velocity relative to the sea and is not as likely to capsize; the release can be effected electromagnetically.

Preliminary results with the accelerometer showed excessive low frequency errors by comparison with a *Doughnut* buoy unit although the frequency spectra computed from simultaneously logged records agreed substantially at frequencies above 0·5 rad/s. These errors are thought to have been due to moisture, since the accelerometer design did not include the shaft seal at the time and it had recently been flooded in an unexpected storm. Experiments in an environmental chamber later showed that condensation troubles can be lessened appreciably by the use of an anti-tracking spray coating on the electrode board and by the introduction of proper sealing. The development of the accelerometer and telemetry equipment is continuing.

7. References

1. Cartwright, D. E., and Smith, N. D., 'Buoy techniques for obtaining directional wave spectra'. 'Buoy Technology', pp. 112–121. (Marine Technology Society, Washington, 1964.)
2. Spiess, F. N., 'Acceleration measurements in the deep ocean', *J. Marine Research*, **16**, No. 1, pp. 23–33, 1957.
3. Tucker, M. J., 'The accuracy of wave measurements made with vertical accelerometers', *Deep-Sea Research*, **5**, pp. 185–92, 1959.

Manuscript first received by the Institution on 8th June 1970 and in final form on 3rd March 1971. (Paper No. 1392/AMMS 38.)

STANDARD FREQUENCY TRANSMISSIONS—June 1971

(Communication from the National Physical Laboratory)

June 1971	Deviation from nominal frequency in parts in 10^{10} (24-hour mean centred on 0300 UT)			Relative phase readings in microseconds N.P.L.—Station (Readings at 1500 UT)	
	GBR 16 kHz	MSF 60 kHz	Droitwich 200 kHz	*GBR 16 kHz	†MSF 60kHz
1	−299·9	0	+0·2	581	594·0
2	−300·0	+0·1	+0·1	581	592·6
3	−299·8	+0·1	+0·2	579	592·0
4	−300·0	+0·1	+0·1	579	591·5
5	−299·9	0	+0·1	578	591·4
6	−300·0	0	+0·1	578	591·7
7	−300·0	0	+0·1	578	592·0
8	−299·9	+0·1	+0·1	577	590·6
9	−299·9	0	+0·1	576	590·8
10	−300·0	+0·1	+0·1	576	589·4
11	−300·0	−0·1	+0·1	576	590·5
12	−299·9	+0·1	+0·1	575	589·8
13	−299·9	+0·1	+0·1	574	588·9
14	−299·9	+0·1	+0·1	573	587·7
15	−299·9	−0·1	+0·1	572	587·8
16	−300·0	0	+0·1	571	587·4
17	−299·9	0	+0·1	571	587·0
18	−299·9	+0·1	+0·1	570	586·5
19	−299·9	+0·1	+0·1	569	585·8
20	−299·9	0	+0·1	568	585·8
21	−299·9	0	+0·1	567	584·9
22	−299·9	+0·1	+0·1	566	583·6
23	−299·9	+0·1	−0·1	565	583·1
24	−299·9	0	−0·2	564	583·1
25	−299·9	+0·1	−0·2	563	582·1
26	−300·0	+0·1	−0·2	563	580·8
27	−299·8	+0·1	−0·2	561	579·8
28	−300·0	+0·1	−0·2	561	579·1
29	−299·9	+0·1	−0·2	560	578·1
30	−299·8	+0·1	+0·1	558	577·5

All measurements in terms of H.P. Caesium Standard No. 334. which agrees with the N.P.L. Caesium Standard to 1 part in 10^{11}.

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

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

U.D.C. 621.313.84–523.2

Analysis and Control of the Permanent-Magnet Stepper Motor

By

P. J. ELLIS,
Ph.D., C.Eng., M.I.E.R.E.†

A stepper motor has the ability to translate digital signals into angular shaft positions. This paper deals with the dynamic characteristics of a 90° permanent magnet motor, when used to switch a load through a sequence of angular steps. The non-linear system equations are derived, and linearized versions of these equations are used to estimate the dynamic response for single-step operation. Multi-step performance is computed directly from the non-linear equations. The effect on performance of a constant current or constant voltage supply is considered, and it is shown that the transient response can be modified with a negative impedance converter.

Notation

T	torque developed by rotor
M	magnetic dipole moment
B	magnetic flux density
K_τ	motor torque constant
K_E	constant relating induced e.m.f. to angular velocity of rotor
J_M	moment of inertia of motor
J_P	moment of inertia of pinion
J_L	moment of inertia of load
D_M	motor damping constant
D_L	load damping constant
D	total damping constant
T_V	frictional torque constant
θ	rotor angular position
θ_O	rotor position at an operating point
ϕ	angular increment of load
i	instantaneous stator winding current
i_O	current at an operating point
I	stator winding supply current
e	stator input voltage
E	stator supply voltage
R	stator winding resistance per phase
R_E	effective resistance of in-phase winding
R_Q	effective resistance of quadrature winding
R_N	negative resistance
Z_{IN}	input impedance of n.i.c.
L	stator inductance per phase
θ_R	rotor starting position
n	gear reduction ratio
ω_n	natural frequency of oscillation
ω_r	damped angular frequency
v	instantaneous rotor velocity
ζ	damping ratio
t	time
Δ	change in a variable
s	Laplace operator.

† Department of Physics, Heriot-Watt University, Edinburgh; formerly of the J. J. Thomson Physical Laboratory, University of Reading.

1. Introduction

The use of electro-mechanical stepper motors to control the angular position of shafts is well established.

Due to their small size and low mean-power requirements, permanent-magnet stepper motors have found applications in satellite-mounted instruments. The selective chopper radiometer[1], now in orbit aboard the *Nimbus IV* weather satellite, uses eight stepper motors to re-position optical filters and mirrors on command.

The successor to the *Nimbus IV* radiometer is due to be launched on the *Nimbus E* spacecraft in 1972. Among the electro-mechanical devices in this instrument is a rotary switch, which inserts infra-red interference filters sequentially into an optical viewing system. The basic problem is to rotate this filter wheel rapidly through a precisely defined angle, and bring it completely to rest within a limited time period.

The necessity of achieving a dynamic response which meets similar requirements is a problem frequently encountered during the design of electro-mechanical systems.

A solution is described which makes use of a 90° permanent-magnet stepper motor as a motive device. Bailey,[2] O'Donouhue,[3] Kieburtz,[4] and Robinson and Taft[5] have analysed the characteristics of this type of motor, and a similar analysis is presented here as an essential basis for an understanding of the system behaviour.

The differential equations which define the response of the motor are first derived for both a constant current and constant voltage supply. Linearized versions of these equations are used to obtain values of the natural frequency and damping ratio, which can be assigned to any arbitrary operating point. Single step response is estimated from the results so obtained, and plots of multi-step operation are generated directly by numerical computation, using the non-linear equations. It is shown that the effective damping ratio may be electronically modified, with the aid of a simple negative impedance circuit.

2. The 90° Permanent-magnet Stepper Motor

2.1 *Static Characteristics*

The motor may be represented by the schematic diagram of Fig. 1(a). The simplest form of rotor is a cylindrical permanent magnet, magnetized across a diameter. The stator consists of two pairs of poles at right angles, wound with identical centre-tapped coils. If winding ABC is energized, a magnetic field is created, aligned with the X-axis. Similarly, a current in winding DEF produces a field aligned with the Y-axis. Interaction of the stator field with that due to the rotor generates a resultant torque, which tends to align the rotor with the stator field.

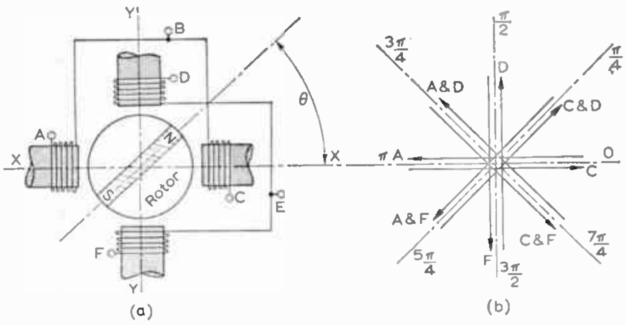


Fig. 1. (a) Motor schematic diagram.
(b) Rotor positions with one or both phases energized.

The fundamental torque equation is given by

$$\mathbf{T} = \mathbf{M} \times \mathbf{B}$$

where **M** is the magnetic dipole moment of the rotor, and **B** is the resultant stator flux density. Thus if winding ABC only is energized, the torque T_x, produced by the flux density B_x is given by

$$T_x = B_x M \sin\theta$$

Similarly, when winding DEF only is energized, the torque T_y is given by

$$T_y = B_y M \cos\theta$$

In this analysis, the stator flux density B is assumed to be a linear function of winding current i, since the air gaps in the magnetic circuit have a linearizing effect on the $B-H$ characteristic. A further assumption is that the X and Y stator windings are identical, thus in each case B is related to i by the same constant k.

The expression $B = ki$ can be combined with $T_x = MB_x \sin\theta$ to give:

$$T_x = K_\tau i_x \sin\theta$$

and

$$T_y = K_\tau i_y \cos\theta$$

where $K_\tau = kM$, the motor torque constant, and i is the current in one 'phase' of either winding. 'Phase' is defined here as that part of a winding between a line and the centre-tap.

2.2 *Energizing Sequence*

It is usually convenient to treat the centre-tap of each winding as a common connexion. Four switches only are then needed to drive the motor, and Fig. 1(b) shows the angular positions which may be defined with one or both coils energized. If a current i flows in one phase of each winding, such that $T_x = K_\tau i \sin\theta$ and $T_y = K_\tau i \cos\theta$, the resultant torque T_r is then given by:

$$T_r = \sqrt{2} K_\tau i \sin\left(\theta + \frac{\pi}{4}\right)$$

Thus with both windings energized, the restoring torque is $\sqrt{2}$ times greater than the value for single coil operation. However, this is accompanied by a reduction in operating efficiency.

In the remainder of this paper, single coil operation only is considered, in which four 90° positions are uniquely defined by the status of the input switches.

2.3 *Dynamic Characteristics—Single Step Operation*

The time-varying functions which have to be considered are the input voltage $e(t)$ and winding current $i(t)$, the rotor position $\theta(t)$, and their time derivatives.

If one phase only is energized, the following differential equations may be used to describe the mechanical and electrical time responses.

The torque equation is

$$J_M \frac{d^2\theta(t)}{dt^2} + D_M \frac{d\theta(t)}{dt} + T_V(t) + K_\tau i(t) \sin\theta(t) = 0 \qquad \ldots\ldots(1)$$

where J_M = moment of inertia of motor and shaft,
D_M = viscous damping coefficient,

and T_V is a function which takes into account static and coulomb friction, and any effects due to unbalance in a gravitational field.

Since the rotor is balanced, and has neither brushes nor slip rings, the value of T_V is determined by the bearing friction alone. In the remaining discussion, T_V is assumed to be small enough to be neglected.

The stator winding equation, neglecting winding capacitance is

$$e(t) = Ri(t) + L\frac{di(t)}{dt} - K_E \frac{d\theta(t)}{dt} \sin\theta(t) \qquad \ldots\ldots(2)$$

where R = stator winding resistance per phase,
L = stator winding inductance per phase,

and K_E is a coefficient which relates the e.m.f. induced in the stator winding to the angular velocity of the rotor.

3. Modes of Operation of a Motor-driven System

Two types of electrical supply are commonly used to power the stator windings. The first type has a high source impedance and approximates to a constant current supply. The second type uses a low source impedance or constant voltage supply. The effect that each of these has on the dynamic response can be deduced from the modified system equations.

3.1 *Constant Current Supply*

The system performance can be quite accurately represented by a modified form of equation (1):

$$J \frac{d^2\theta(t)}{dt^2} + D \frac{d\theta(t)}{dt} + K_\tau I \sin\theta(t) = 0 \qquad \ldots\ldots(3)$$

where J = total inertia of motor and load,

D = total damping coefficient,

and I = stator input current per phase.

It is seen that the stator winding resistance, inductance, and generated e.m.f. do not appear in this equation, and it follows that the effects due to these quantities may be minimized by the use of high impedance supply. In practice, however, the value of coil inductance will determine the maximum rate at which current switching may take place.

More information about the system response under constant current conditions may be gained by linearizing the basic equation about an arbitrary operating point θ_O. Thus:

$$J\frac{d^2\Delta\theta(t)}{dt^2}+D\frac{d\Delta\theta(t)}{dt}+K_\tau I\cos\theta_O\Delta\theta(t)=0 \qquad \text{......(4)}$$

where $\Delta\theta(t)$ is a small increment from the operating point θ_O.

The Laplace transform of this expression is now given, with initial conditions $\theta = -\theta_R$, $d\theta/dt = 0$, i.e.

$$Js^2\Delta\theta(s)+Js\theta_R+Ds\Delta\theta(s)+D\theta_R+K_\tau I\cos\theta_O\Delta\theta(s)=0$$

from which

$$\Delta\theta(s)=\frac{-\dfrac{\theta_R}{J}\left(s+\dfrac{D}{J}\right)}{s^2+\dfrac{Ds}{J}+\dfrac{K_\tau I}{J}\cos\theta_O} \qquad \text{......(5)}$$

The undamped angular frequency ω_n and damping ratio ζ can be derived from this expression, and a value of ω_n and ζ can be assigned to any operating point θ_O. Thus

$$\omega_n=\sqrt{\frac{K_\tau I\cos\theta_O}{J}} \qquad \text{......(6)}$$

and

$$\zeta=\frac{D}{2\sqrt{JK_\tau I\cos\theta_O}} \qquad \text{......(7)}$$

When $\theta_O = 0$, the motor is in the 'detent' position, and the natural frequency for small oscillations about this point is given by:

$$\omega_n=\sqrt{\frac{K_\tau I}{J}} \qquad \text{......(8)}$$

When $\theta_O = 45°$

$$\omega_n=\sqrt{\frac{K_\tau I}{J\sqrt{2}}}$$

and

$$\frac{\omega_n(0)}{\omega_n(45°)}=0{\cdot}84$$

The angular frequency ω_r of the damped oscillations about the 'detent' position is:

$$\omega_r=\sqrt{\frac{K_\tau I}{J}-\left(\frac{D}{2J}\right)^2} \qquad \text{......(9)}$$

From the foregoing expressions, the rate of decay of oscillatory transients may be estimated, but it is more difficult to determine the time taken for the motor to switch through one positional increment of $\pi/2$ radians (90°).

If the damping term is small, i.e. ζ is less than (say) 0·2, an approximate value for the switching time may be derived by considering the undamped non-linear equation:

$$\frac{d^2\theta(t)}{dt^2}=-\frac{K_\tau I}{J}\sin\theta(t)$$

Now if the instantaneous angular velocity of the rotor is

$$v=\frac{d\theta(t)}{dt}$$

then

$$\frac{d^2\theta(t)}{dt^2}=v.\frac{dv}{d\theta(t)}$$

and therefore

$$\int v.dv=-\frac{K_\tau I}{J}\int\sin\theta(t)\,d\theta(t)$$

and

$$v=\sqrt{\frac{2K_\tau I}{J}\cos\theta(t)}$$

If the instantaneous velocity is now integrated over the range $\theta(t) = -90°$ to $\theta(t) = 0$, an expression for the mean velocity can be obtained.

Thus

$$v_{\text{average}}=\frac{2}{\pi}\int_0^{\pi/2}\sqrt{\frac{2K_\tau I}{J}}\sqrt{\cos\theta(t)}\,d\theta(t)$$

A numerical integration gives the result

$$v_{\text{average}}=0{\cdot}76\sqrt{\frac{2K_\tau I}{J}} \qquad \text{......(10)}$$

assuming that $v = 0$ when $\theta(t) = 90°$.

This analysis enables the single-step response of a motor driven system to be estimated under constant current supply conditions. By single-step operation, it is meant that all transients are allowed to decay essentially to zero before the next driving pulse is initiated.

Multi-step operation is more difficult to analyse, since the initial conditions for one step are determined by the conditions at the end of the preceding step. Robinson and Taft[5] make use of a graphical technique, phase plane analysis,[6] to derive the multi-step response from the non-linear equation of motion.

A digital computer may also be used to solve the equations for specific coefficient values and a known switching sequence. It is convenient to obtain the output as an automatic plot of position and velocity against time. Some results with typical parameters are shown in Figs 2 and 3.

3.2 *Constant Voltage Supply*

Operation from a constant voltage source is considered for two reasons: (a) a relatively simple circuit is required to switch the windings across a stabilized supply; (b) the

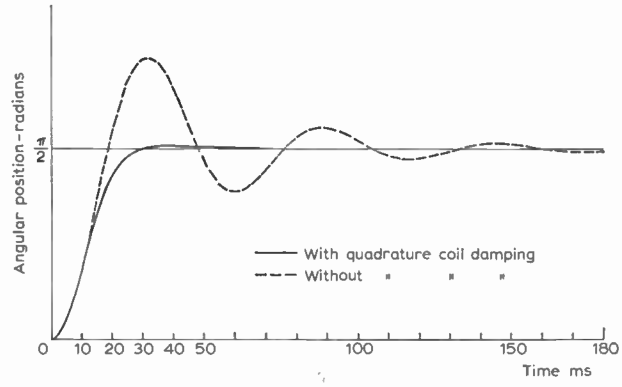


Fig. 2. (a) Single step position response, computed.

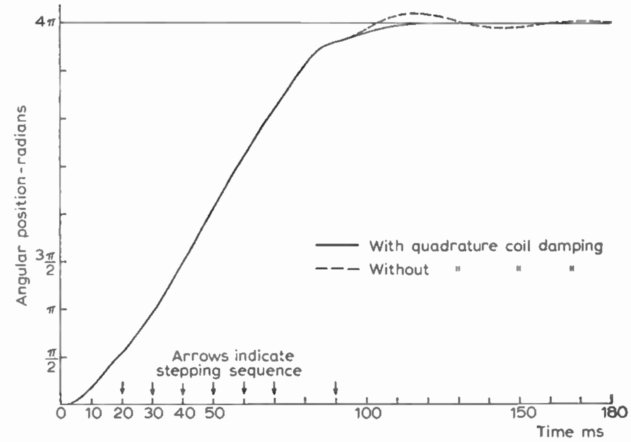


Fig. 3. (a) Multi-step position response, computed.

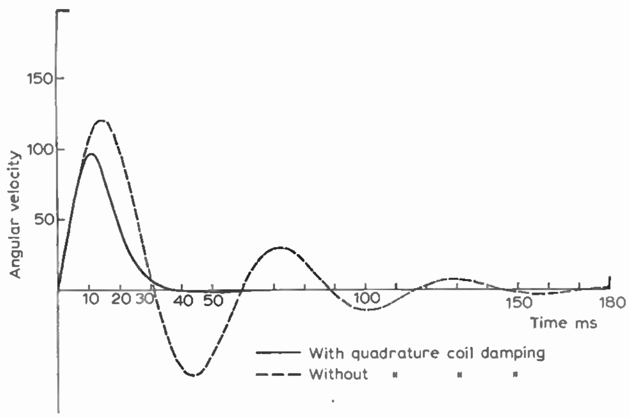


Fig. 2. (b) Single step velocity response, computed.

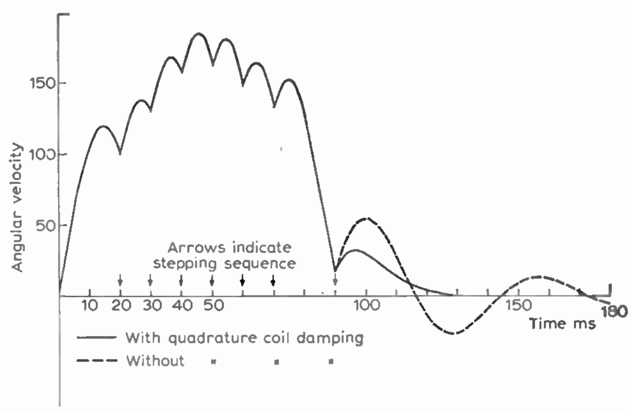


Fig. 3. (b) Multi-step velocity response, computed.

analysis points the way to obtaining a modified response by the use of an external electronic circuit.

Use is made of both the original non-linear equations:

$$J\frac{d^2\theta(t)}{dt^2}+D\frac{d\theta(t)}{dt}+K_\tau i(t)\sin\theta(t)=0 \quad \ldots\ldots(1)$$

where $i(t)$ is a function of $e(t)$ and $\theta(t)$,

and

$$e(t)=Ri(t)+L\frac{di(t)}{dt}-K_E\frac{d\theta(t)}{dt}\sin\theta(t) \quad \ldots\ldots(2)$$

As in the previous case, these equations may be linearized for small increments $\Delta\theta(t)$ and $\Delta i(t)$ about the arbitrary operating points θ_O and i_O.

Thus:

$$J\frac{d^2\Delta\theta(t)}{dt^2}+D\frac{d\Delta\theta(t)}{dt}+K_\tau i_O\cos\theta_O\Delta\theta(t)+ +K_\tau\sin\theta_O\Delta i(t)=0 \quad \ldots\ldots(11)$$

and

$$\Delta e(t)=R\Delta i(t)+L\frac{d\Delta i(t)}{dt}-K_E\sin\theta_O\frac{d\Delta\theta(t)}{dt} \quad \ldots\ldots(12)$$

i_O is a constant, having the value E/R, where E is the supply voltage, and R is the winding resistance per phase.

The time-varying quantities can now be transformed, giving algebraic equations in the complex frequency plane. The initial conditions are as follows:

$$\theta=-\theta_R,\qquad \frac{d\theta(t)}{dt}=0,\qquad i_o=0,$$

and a step voltage E is applied at $t=0$.

Then:

$$Js^2\Delta\theta(s)+Js\theta_R+Ds\Delta\theta(s)+D\theta_R+\frac{K_\tau E}{R}\cos\theta_O\Delta\theta(s)+ +K_\tau\sin\theta_O\Delta i(s)=0 \quad \ldots\ldots(13)$$

and

$$\frac{E}{S}=\Delta i(s)(R+Ls)-sK_E\sin\theta_O\Delta\theta(s)-K_E\sin\theta_O\theta_R. \quad \ldots\ldots(14)$$

From (14)

$$\Delta i(s)=\frac{E+s^2K_E\sin\theta_O\Delta\theta(s)+sK_E\sin\theta_O\theta_R}{s(R+Ls)}$$

Combining and solving for $\Delta\theta(s)$:

$$\Delta\theta(s) = \frac{-\dfrac{1}{JL}\,[K_\tau E \sin\theta_O + sK_\tau K_E \sin^2\theta_O\theta_R + \theta_R s(D+Js)(R+Ls)]}{s\left[s^3 + s^2\left(\dfrac{R}{L}+\dfrac{D}{J}\right) + s\left(\dfrac{DR}{JL}+\dfrac{K_\tau E}{JR}\cos\theta_O + \dfrac{K_\tau K_E}{JL}\sin^2\theta_O\right) + \dfrac{K_\tau E}{JL}\cos\theta_O\right]} \quad \text{......(15)}$$

In order to find the time response at any point θ_O, the roots of the characteristic equation must be evaluated. The roots of this cubic equation can be numerically computed for any specific set of coefficients.

In many practical cases, the electrical time-constant L/R has a much smaller value than the mechanical time-constant J/D. The effect of a finite value of L/R on the roots of the equation can be deduced by letting $L \to 0$, and comparing the roots so obtained with the roots of the cubic equation.

If the numerator and denominator of (15) are multiplied by L/R, and then $L \to 0$, the characteristic equation becomes:

$$s\left[s^2 + s\left(\frac{D}{J} + \frac{K_\tau K_E}{JR}\sin^2\theta_O\right) + \frac{K_\tau E}{JR}\cos\theta_O\right] = 0 \quad \text{......(16)}$$

giving:

$$\omega_n = \sqrt{\frac{K_\tau E}{JR}\cos\theta_O} \quad \text{......(17)}$$

and

$$\zeta = \frac{D + \dfrac{K_\tau K_E}{R}\sin^2\theta_O}{2\sqrt{\dfrac{JK_\tau E}{R}\cos\theta_O}} \quad \text{......(18)}$$

....

Table 1 is a comparison of the roots of (15) over a range of θ_O, for values of L/R between 0 and 0·1 J/D. The other quantities in these expressions are derived from the typical system described in Section 7. Case (b) uses the measured value of L/R. The location of these roots in the s-plane completely determines the transient response, and it is seen that the three principal roots are only slightly modified as L/R increases from 0 to 0·1 J/D. The high negative value of S_3 ensures that it has little effect on the final result. Thus the relations (17) and (18) may be used to describe the system at any point θ_O, and the single step response may be approximately determined, as in the case of constant-current drive.

The expressions derived in this section enable the dynamic response with a constant voltage supply to be approximately determined for single step operation. These calculations are simplified if $L/R < 0{\cdot}1\ J/D$. Examination of the damping ratio (18), shows an additional term $K_\tau K_E/R \sin^2\theta_O$, which is due to a current induced in the energized winding by the rotating magnet.

In Section 6, it is shown that a similar term in the quadrature winding may be electronically modified to give a critically-damped response.

Table 1. Poles of transfer function, linearized about an operating point θ_O. Showing effect of electrical time-constant L/R

		$\theta_O = 0°$	$\theta_O = 30°$	$\theta_O = 45°$	$\theta_O = 60°$	$\theta_O = 75°$	$\theta_O = 85°$
(a) $\frac{L}{R} = 0$	S_1, S_2	$-17 \pm j112{\cdot}5$	$-19{\cdot}08 \pm j104$	$-21{\cdot}1 \pm j93{\cdot}5$	$-23{\cdot}1 \pm j77{\cdot}1$	$-24{\cdot}6 \pm j52{\cdot}5$	$-25{\cdot}1 \pm j22{\cdot}5$
	S_4	0	0	0	0	0	0
(b) $\frac{L}{R} = 0{\cdot}014\,\frac{J}{D}$	S_1, S_2	$-17{\cdot}03 \pm j112{\cdot}7$	$-19{\cdot}01 \pm j104{\cdot}3$	$-21{\cdot}1 \pm j93{\cdot}5$	$-23{\cdot}25 \pm j77{\cdot}4$	$-24{\cdot}71 \pm j52{\cdot}7$	$-25{\cdot}3 \pm j22{\cdot}4$
	S_3	$-2{\cdot}496 \times 10^3$	$-2{\cdot}492 \times 10^3$	$-2{\cdot}488 \times 10^3$	$-2{\cdot}484 \times 10^3$	$-2{\cdot}481 \times 10^3$	$-2{\cdot}478 \times 10^3$
	S_4	0	0	0	0	0	0
(c) $\frac{L}{R} = 0{\cdot}05\,\frac{J}{D}$	S_1, S_2	$-17{\cdot}05 \pm j112{\cdot}7$	$-19{\cdot}21 \pm j104{\cdot}6$	$-21{\cdot}27 \pm j94{\cdot}0$	$-23{\cdot}41 \pm j77{\cdot}9$	$-25{\cdot}17 \pm j53{\cdot}0$	$-25{\cdot}6 \pm j22{\cdot}4$
	S_3	$-6{\cdot}809 \times 10^2$	$-6{\cdot}766 \times 10^2$	$-6{\cdot}725 \times 10^2$	$-6{\cdot}682 \times 10^2$	$-6{\cdot}647 \times 10^2$	$-6{\cdot}637 \times 10^2$
	S_4	0	0	0	0	0	0
(d) $\frac{L}{R} = 0{\cdot}1\,\frac{J}{D}$	S_1, S_2	$-17{\cdot}06 \pm j112{\cdot}6$	$-19{\cdot}19 \pm j104{\cdot}9$	$-21{\cdot}34 \pm j94{\cdot}7$	$-23{\cdot}74 \pm j78{\cdot}6$	$-25{\cdot}6 \pm j53{\cdot}5$	$-26{\cdot}4 \pm j22{\cdot}4$
	S_3	$-3{\cdot}409 \times 10^2$	$-3{\cdot}366 \times 10^2$	$-3{\cdot}323 \times 10^2$	$-3{\cdot}275 \times 10^2$	$-3{\cdot}238 \times 10^2$	$-3{\cdot}222 \times 10^2$
	S_4	0	0	0	0	0	0

4. The Effect of Gears

The foregoing arguments assume that the system has a total inertia $J = (J_M + J_L)$, where J_L is the inertia of the load, and damping coefficient $D = (D_M + D_L)$, where D_L is the load damping coefficient. This implies that the load is coupled directly to the motor shaft. A gear reduction between motor and load will materially alter the system characteristics, and it is instructive at this stage to consider the effect of such an introduction. A gear ratio of n : 1 between motor and load will have the following consequences:

1. The 'reflected' value of load inertia referred to the motor shaft is J_L/n^2. Thus the total effective inertia is $(J_M + J_P + J_L/n^2)$, where J_P is the inertia of the driving pinion.

2. Since the restoring torque is low for small angles around the detent position, errors in pointing accuracy will occur due to the effect of bearing friction. If the error in motor shaft position is $\Delta\theta$, the error at the load is $\Delta\theta/n$, neglecting backlash.

3. The torque available to drive the load is equal to n times the motor output torque. Thus any friction effects associated with the load bearings are reduced by the factor n.

4. The maximum amplitude of an oscillatory transient for a 90° motor is $\pm\pi/2$ radians. The corresponding load transient is limited to $\pm\pi/2n$.

5. Examination of equations (7) and (18) show that the damping ratio ζ is increased as the effective inertia is reduced. This will slow down the speed of response for a given angular load increment, but will also increase the rate of decay of the transient terms.

5. Multi-step Operation

It is generally required to rotate the load through an angle and stop it within a specified time. This usually involves pulsing the motor through a number of steps. If a fast response is required, the motor does not have time to settle down in one 'detent' position before the next pulse is applied. Thus the initial conditions for each pulse will depend on the effect of all the preceding pulses on the rotor. The overall dynamic response can be numerically computed for any arbitrary pulse sequence, using a computer program which solves the original non-linear equations.

For instance, assume that the stator inductance has a negligible effect on the response. Then the value of $i(t)$, defined by equation (2), can be inserted directly into equation (1) to give:

$$J\frac{d^2\theta(t)}{dt^2} + D\frac{d\theta(t)}{dt} + \frac{K_\tau e(t)}{R}\sin\theta(t) + \frac{K_\tau K_E}{R}\sin^2\theta(t)\frac{d\theta(t)}{dt} = 0 \qquad \ldots\ldots(19)$$

where $e(t)$ is the applied step voltage.

By a change of variables, the expression may be reduced to two simultaneous first-order equations. These can then be inserted into a standard procedure for solving differential equations, such as that due to P. Naur.[7] The program may be extended to produce automatic plots of angular position and velocity versus time. Figure 3 shows an example of the plots which simulate the response of the satellite filter switch.

6. The Electronically Damped System

With a constant voltage supply, it is seen that the moving rotor induces a current in the energized stator winding. The magnitude of this current is given by:

$$\frac{K_E}{R}\sin\theta(t)\frac{d\theta(t)}{dt}$$

The winding is essentially short-circuited for transient induced voltages and currents. A similar effect can be produced by short-circuiting any of the unused windings during one pulse period. If a quadrature winding, such as winding 2 in Fig. 1, is short-circuited, the current term becomes:

$$\frac{K_E}{R}\cos\theta(t)\frac{d\theta(t)}{dt}$$

It follows that even with a constant current drive circuit, the unused windings may be sequentially shorted to increase the effective damping ratio. Generally, however, the total damping ratio still falls far short of unity. The ideal situation is one in which the system has an inherently low damping ratio to give a fast response during most of the stepping sequence, and is critically damped during the final one or two steps.

It is possible to increase the damping ratio by inserting a negative resistance in series with the winding resistance R. If the effective resistance of the combination is R_E, then for a constant voltage supply, equation (18) becomes:

$$\zeta = \left\{\frac{D + \dfrac{K_\tau K_E}{R_E}\sin^2\theta_O}{2\sqrt{\dfrac{JK_\tau E}{R_E}\cos\theta_O}}\right\} \qquad \ldots\ldots(20)$$

R_E in this equation refers, of course, to the resistance of the energized winding. If the negative resistance is connected across the quadrature winding, giving an effective resistance R_Q, then the damping term becomes:

$$\zeta_{cv} = \left\{\frac{D + \dfrac{K_\tau K_E}{R}\sin^2\theta_O + \dfrac{K_\tau K_E}{R_Q}\cos^2\theta_O}{2\sqrt{\dfrac{JK_\tau E}{R}\cos\theta_O}}\right\} \qquad \ldots\ldots(21)$$

for a constant voltage supply,
and

$$\zeta_{cc} = \left\{\frac{D + \dfrac{K_\tau K_E}{R_Q}\cos^2\theta_O}{2\sqrt{\dfrac{JK_\tau E}{R}\cos\theta_O}}\right\} \qquad \ldots\ldots(22)$$

for a constant current supply.

A negative resistance circuit suitable for switching across one winding is shown in Fig. 4(a). Figure 4(b)

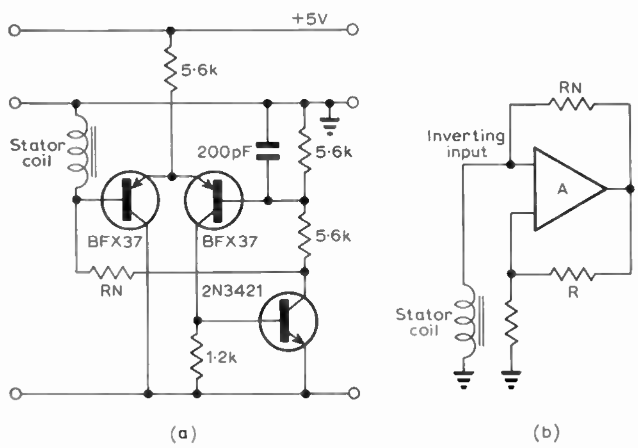


Fig. 4. (a) Negative impedance circuit. (b) Schematic.

is a schematic from which the input impedance Z_{IN} may be calculated.

For a difference amplifier with a high impedance input

$$Z_{IN} = -R_N \left(\frac{A-2}{A+2}\right)$$

and when $A \gg 1$ then $Z_{IN} \simeq -R_N$.

7. Performance of a Typical System

The rotary switch in the satellite instrument consists of a size eleven stepper motor with an 8 : 1 gear reduction between motor and load. The fixed parameters for this system are:

$$\begin{aligned} J\ (\text{effective}) &= 5{\cdot}6\times10^{-7}\ \ \text{kg m}^2\\ D &= 1{\cdot}91\times10^{-5}\ \text{N m s/rad}\\ K_\tau &= 4{\cdot}54\times10^{-2}\ \text{N m/A}\\ K_E &= 3{\cdot}0\times10^{-2}\ \ \text{V s/rad}\\ E &= 24\ \text{V}\\ R &= 150\ \Omega\\ L &= 60\times10^{-3}\ \text{H}\\ R_N &= -135\ \Omega \end{aligned}$$

Figure 2 shows the single-step computed response, with and without electronic damping. Figure 3 shows the computed multi-step response for eight steps. The value of R_Q may be estimated from (21) in the constant voltage case, or (22) in the constant current case.

Critical damping at $\theta_O = 0$, occurs when

$$R_Q = \frac{K_\tau K_E}{\left\{2\sqrt{\dfrac{JK_\tau E}{R}} - D\right\}}$$

and $R_N = (R_Q - R)$.

Figure 5 was obtained by direct measurement on the prototype satellite instrument, using an optical position sensor mounted on the motor shaft. The position output versus time is plotted on the same scale as Fig. 3(a).

8. Conclusions

It has been established that approximate expressions may be used to describe the single-step performance of a permanent-magnet stepper motor, with either a constant current or constant voltage supply. Numerical methods have the advantage when considering multi-step operation, and the graphical results of these computations are shown for a typical system, and compared with measurements.

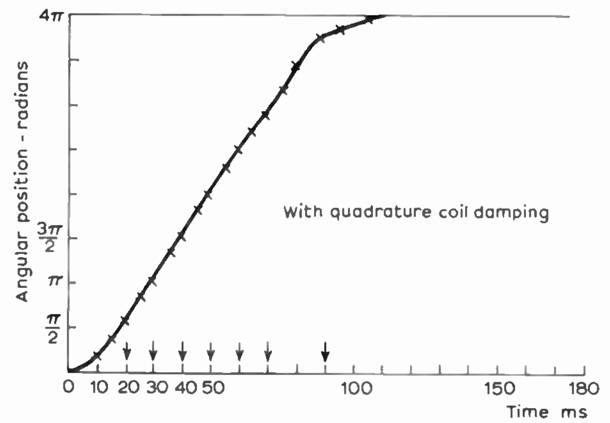


Fig. 5. Multi-step position response, measured.

Two conflicting requirements tend to arise when using stepper motors. For any given design, the most rapid response is achieved if the damping ratio is much less than unity. However, a decaying oscillation about the final 'detent' position is an inevitable consequence of this condition, and the settling time may be too long. The electronic damping circuit described can be switched into the appropriate stator winding for the duration of the terminating pulse. Thus for multi-step operation, it is possible to combine a fast response with any desired degree of damping at the conclusion of the pulse train.

9. Acknowledgments

The work described is part of the *Nimbus* Satellite Radiometer Programme, supported by the Science Research Council. The author is indebted to Mr. H. Hadley and staff at the Rutherford High Energy Laboratory, for carrying out the experimental work. Thanks are also due to Dr. G. Peckham of Reading University, and Professor S. D. Smith and Dr. M. Cross of Heriot-Watt University, Edinburgh, for their help and encouragement.

10. References

1. Abel, P. G., *et al.*, 'Remote sounding of atmospheric temperature from satellites'. II—'The selective chopper radiometer for *Nimbus D*', *Proc. Royal Society Lond.*, A, **320**, pp. 35–55, 1970.
2. Bailey, S. J., 'Incremental servos. Part 1—Stepping vs. stepless control', *Control Engrg*, **7**, pp. 123–7, November 1960.
3. O'Donouhue, J. P., 'Transfer function for a stepper motor', *Control Engrg*, **8**, pp. 103–4, November 1961.
4. Kieburtz, R. B., 'The step motor—the next advance in control systems', *I.E.E.E. Trans. Automatic Control*, **AC-9**, pp. 98–104, January 1964.
5. Robinson, D. J. and Taft, C. K., 'A dynamic analysis of magnetic stepping motors', *I.E.E.E. Trans. Industrial Electronics and Control Instrumentation*, **IECI-16**, No. 2, pp. 111–25, September 1969.
6. Thaler, G. J. and Brown, R. G., 'Analysis and Design of Feedback Control Systems'. (McGraw-Hill, New York, 1960.)
7. Naur, P., 'Algorithm 9, Runge-Kutta integration', Collected Algorithms from C.A.C.M., May 1960.

Manuscript first received by the Institution on 27th October 1970 and in final form on 26th January 1971. (Paper No. 1393/IC46.)

U.D.C. 621.317: 621.385.63: 621.397.61.029.64

Measuring Techniques for U.H.F. Television Transmitters Employing Klystrons

By
J. F. H. BINNS†

Reprinted from the Proceedings of the Conference on Television Measuring Techniques held in London from 11th to 13th May 1970.

The paper describes the measuring techniques associated with complete u.h.f. transmitter installations, indicating the large range of test equipment that is required. An indication of the considerable distortions produced by a klystron amplifier is given, but no attempt is made to detail the mechanisms involved.

1. Introduction

The parameters which have to be measured fall into three categories:

(a) Output signal levels.

(b) Output signal distortions.

(c) Output signal stability.

2. Measurement of Power Levels

The output powers of u.h.f. klystron transmitters are generally at levels allowing classic calorimetric methods to be used. It is common to dissipate the radio frequency power in the coolant which is circulated through a section of transmission line.

The power dissipated is calculated from the flow rate and temperature rise observed. The frequency-modulated sound output signal presents no problem, but in the case of the vision output signal, the mean power measured with modulation present has little significance.

In order to establish the vision power it is necessary to monitor the vision output via a detector and oscilloscope as well as measuring the power. The detector signal is supplied by a directional coupler mounted in the feeder which is carrying the power to the test load.

The procedure is to drive the klystron with a continuous wave signal at vision carrier frequency to give a power output measured at the test load equal to the power at either the required blanking level or at the synchronizing pulse level. The d.c. shift of the oscilloscope trace produced by the detector output is carefully noted. Modulation is then restored to the transmitter and the video signal feeding the vision modulator is adjusted to place the blanking level or synchronizing pulse level, as displayed on the oscilloscope, at the previously noted level. If there is any doubt about the linearity of the detector the power can be established by c.w. measurement separately at both blanking level and synchronizing pulse level.

Day-to-day measurement of power does not involve the above procedure as the transmitters are furnished with detector-fed meters calibrated in power. The vision power metering is arranged as a peak reading indicator and normally gives a substantially constant indication, irrespective of the picture content of the transmitters.

2.1 *Measurements on the Output Feeder Systems*

The power meters described above are normally arranged to indicate reflected power by switching the meter circuit to a detector which is fed from a directional coupler set to monitor the reflected wave in the feeder. The measurement of both forward and reflected power enables a check to be maintained on the match of the feeder network. It also becomes possible to check if there is any spurious power feed from the sound transmitter back into the vision transmitter or vice versa, a condition which can obtain if the combining network is incorrectly set. This 'cross insertion' is checked by switching off one transmitter; any power then indicated on its reflected power metering must be power from the other transmitter.

The combining network is a four-port network with two inputs and two outputs. The two inputs are from the sound and vision transmitters and the network when correctly set up delivers the combined power to the output port. The fourth port is connected to a 'balancing load' which is also provided with a detector-fed power indicator.

The power indicators on the two inputs, plus the indicator of balancing load power, give sufficient information for routine checking and setting of the combining network resonators.[1]

3. Performance Measurements

There are two categories of performance measurement:

(a) Those made at radio frequency.

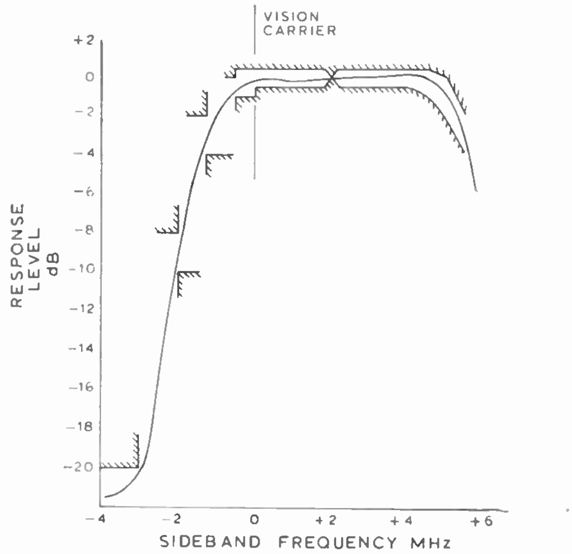
(b) Those made at video or audio frequency after demodulation.

3.1 *Vision Frequency Response*

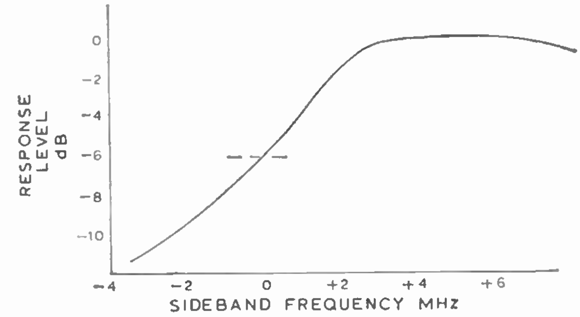
Vision frequency response can be measured by a variety of techniques. It can be measured by observing the video output of a television demodulator and as a routine check this is to be recommended as practically all other vision performance is measured after demodulation and is therefore influenced by this overall measurement.

Where adjustment is required, however, as the demodulated frequency response is the resultant of a vestigial sideband transmitter characteristic and a matched vestigial sideband receiver, it is not advisable to attempt anything but the most minor change. (See Fig. 1.)

† Marconi Communication Systems Ltd., Chelmsford, Essex.



(a) Typical transmitter response.



(b) Demodulator response for vision measurements.

Fig. 1.

The method used when adjustment or initial setting is required is to measure the amplitude of the sidebands produced by modulation at specific frequencies. The sideband amplitudes are measured using a selective voltmeter coupled to a directional coupler in the output feeder.

The instrument requires a very narrow bandwidth in order that modulation frequencies close to the vision carrier may be measured without being influenced by the presence of the large amplitude of the carrier. A typical instrument used for these measurements will allow 100 kHz sidebands to be measured which may be 30 to 40 dB less than the carrier amplitude.

This method can be very accurate and is frequently used as a final check of response in transmitter acceptance tests, but as it is a point-by-point method, it would be laborious to use for initial setting of response.

The swept frequency equivalent of the measurement is normally used and the instrument is termed a 'sideband analyser'. This test equipment is essentially a selective voltmeter with tuning which is swept over the radio frequency range of the transmitter to be checked. In order that the output of the selective voltmeter part of the instrument will display a sideband response of the transmitter, it is necessary for the instrument to produce a swept video frequency signal with which to modulate the transmitter.

The two swept signals have to track extremely accurately such that the video frequency at any instant in time is equal to the frequency difference between vision carrier and the frequency to which the selective receiver is tuned. The tracking problem is particularly acute as the selectivity must be very narrow in order to satisfactorily record the frequency response in the region of vision carrier frequency.

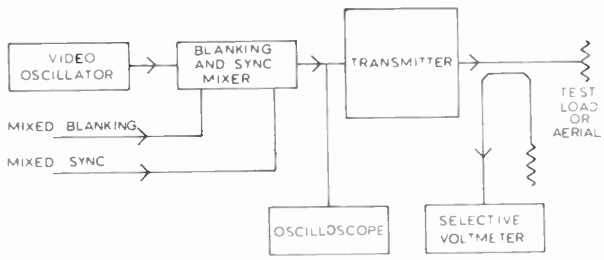
In use, the display oscilloscope can be placed next to the part of the transmitter requiring adjustment and the effect of adjustments observed as they are made.

Monitor couplers are provided such that the sideband response can be traced right through the radio frequency section of the transmitter. Figure 2 shows the test set-up for frequency response measurement.

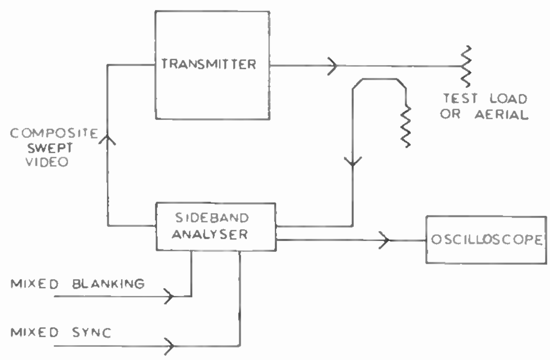
3.2 *Video Measurements after Demodulation*

Once the correct frequency response has been established the remaining performance must be measured using a television demodulator. There are, however, inherent shortcomings in a vestigial sideband system which become apparent when it is necessary to interpose a demodulator in the measuring system.

The basic limitation is due to the fact that in order to produce an effective flat frequency response overall, the demodulator is shaped such that the vision carrier frequency is 6 dB down the slope at one side of the response. This ensures that at the low modulation frequencies where two sidebands exist, the resultant video output is the same as results from the single sideband which occurs for the higher modulation frequencies.



(a) Point-by-point plot.



(b) Sideband analyser.

Fig. 2. Test set-up for sideband response measurement.

Unfortunately, when the signal is detected, a 'quadrature' or single sideband distortion occurs which is dependent on signal amplitude. Some measurements, for instance pulse measurements, have to be made with a small amplitude signal in order to avoid pulse narrowing which is a typical effect of quadrature distortion. (See Fig. 3.)

The specification of the best television demodulators is not very much better than the specification which must be aimed for by the transmitter manufacturer. The difficulty also exists that if a demodulator performance is suspect or is challenged, then an ideal transmitter is required in order to test it.

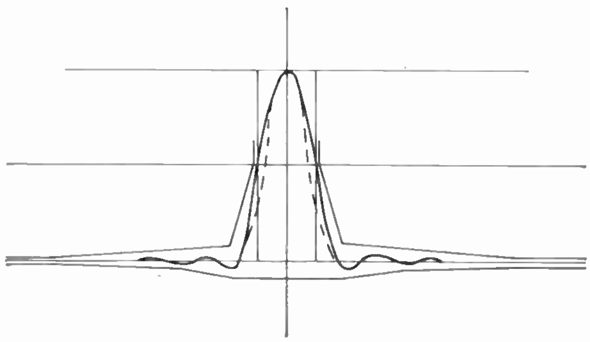


Fig. 3. Typical demodulated $2T$ sine-squared pulses. The dotted line shows a full black-to-white amplitude pulse suffering narrowing due to quadrature distortion. The full line shows a pulse 10 dB reduced in level. (The oscilloscope gain is increased to fill the graticule.)

3.3 *Linearity, Differential Gain and Differential Phase*

These parameters being measured as video signals do not involve any new or unusual technique, but it is worth mentioning, however, that the transmitter engineer is dealing with a somewhat different problem to that which faces, for instance, a studio engineer.

The studio engineer may be concerned with a very tight specification for a video amplifier which may be part of a chain of ten amplifiers. The transmitter engineer is concerned with measuring the resultant of the gross distortion produced in the klystron amplifier and the predistortion of the transmitter correction circuits.

The distortion introduced by a klystron amplifier can be appreciated by the fact that the slope of its transfer characteristic at blanking level may be some 6 dB less than at white level, and at the peak output at the synchronizing pulse tip the transfer characteristic is frequently very nearly horizontal. The transmitter user in the interests of running economy tends to keep the klystron high tension supply to as low a voltage as possible consistent with achieving his used peak power output. (See Fig. 4.)

Thus 6 to 10 dB of differential gain correction is typical and perhaps 10 to 20 degrees of differential phase correction. The correction is normally required to match the distortion such that the transmitted differential gain error does not exceed 0·5 dB and the differential phase error does not exceed $\pm$ 3 degrees relative to a reference at blanking level.

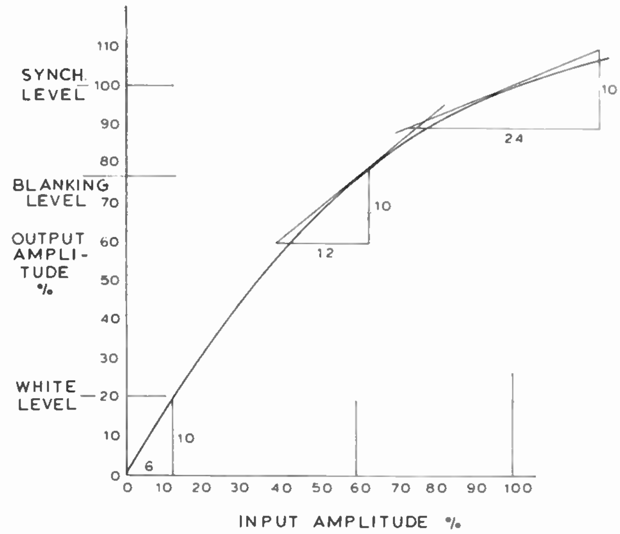
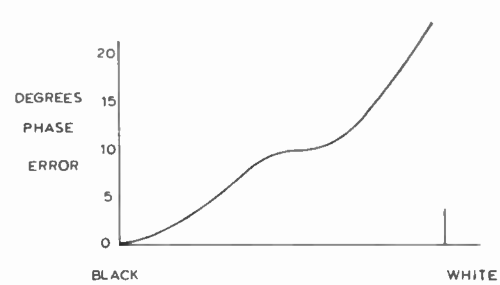


Fig. 4. Typical klystron transfer characteristic showing a saturated output of 110%. Gain at blanking level—6 dB and gain at synchronizing pulse level—12 dB relative to gain at white level.

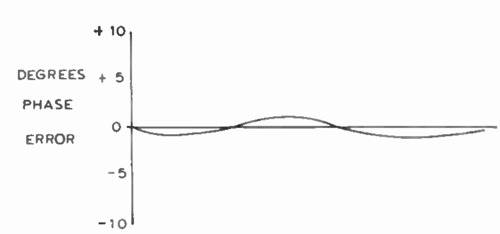
A limitation occurs in that some klystrons exhibit a differential phase distortion characteristic containing an inflexion. In this case the best correction that can be applied leaves the inflexion undisturbed. (See Fig. 5.)

In addition to measuring differential gain and phase a separate measurement of linearity is important in connection with u.h.f. transmitters due to the unfortunate fact that klystron distortions may not be consistent over their bandwidth.

For technical and economic reasons the vestigial shaping is frequently made at low power level before the signal is amplified to output power by the klystron. As previously mentioned, the klystron amplifier has a far from linear transfer characteristic and this results in the lower sidebands, which have been filtered out, being



(a) Differential phase distortion with inflexion as found with some klystrons.



(b) Compromise correction applied.

Fig. 5.

re-constituted by intermodulation. This effect is more pronounced as the power level is increased and shows itself in two measurements. The frequency response changes with power level and the linearity distortion differs from the differential gain distortion. Normally there is more distortion to low modulation frequencies making additional correction circuitry necessary which gives predistortion of the low modulation frequencies, but leaves the higher frequencies, including colour sub-carrier, undisturbed.

The usual method of linearity measurement is to modulate the transmitter with a staircase waveform, the 'risers' of each step having precisely the same amplitude and form. A suitable filter is inserted between the demodulator output and the oscilloscope which leaves a row of spikes displayed corresponding to the riser amplitudes. If the linearity distortion is completely corrected the spikes will all have the same height. Specifications normally require a similar performance to that for differential gain. (See Fig. 6.)

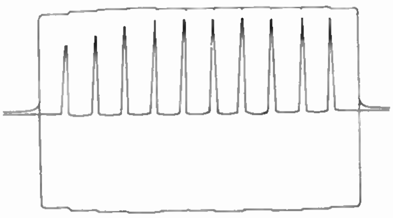


Fig. 6. Typical transmitter differential gain and linearity measurement displays showing corrected differential gain and need for further linearity correction.

An alternative approach to this problem of differing linearity and differential gain is to provide pre-distortion circuitry which operates at radio frequency. Clearly if a radio frequency circuit can be produced which, when cascaded with the klystron, forms a combination with an ideal transfer characteristic at all frequencies within the channel, no distortions will be produced.

It is difficult to design such circuits to function at the frequency of operation of a u.h.f. klystron, but transmitters using intermediate frequency modulation followed by frequency conversion to the radiated channel, successfully employ this technique.

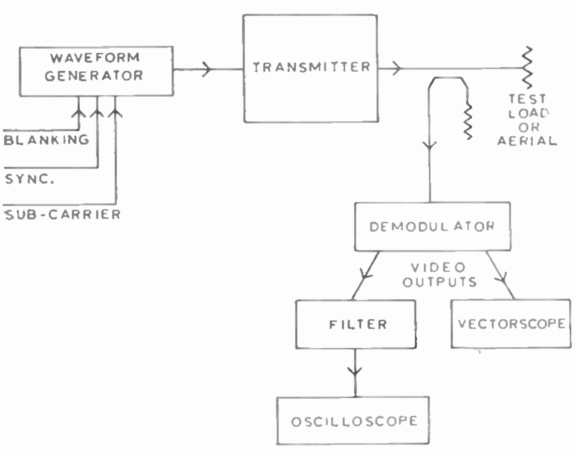


Fig. 7. Test set-up for linearity, differential gain and phase.

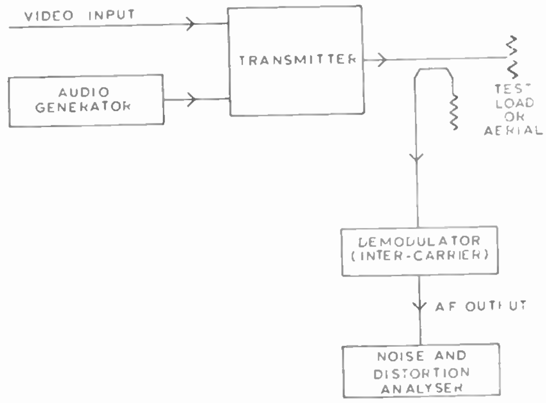


Fig. 8. Test set-up for incidental phase modulation check.

Figure 7 shows the test set-up for linearity, differential gain and phase.

3.4 *Incidental Phase Modulation*

The majority of domestic television receivers and professional demodulators provide a sound output by the 'inter-carrier' system. They make use of the difference between sound and vision carrier frequencies, a frequency which exists at the video detector output if this is fed with both sound and vision frequencies.

This 'inter-carrier' signal is amplified, limited and fed to an f.m. discriminator to provide an audio signal. Clearly such a discriminator will give an audio output if either carrier is frequency modulated.

It is therefore an important requirement that the vision amplitude modulator must not give rise to a spurious frequency modulated signal. This is, in fact, quite difficult to avoid because due to the high modulation frequencies associated with video modulation, the vision carrier only requires to be phase modulated by a few degrees to give rise to an appreciable frequency modulated signal. In particular, the high depth of amplitude modulation at video line rate can easily produce a line rate signal on the audio output of a receiver if incidental phase modulation exists in the transmitter. Although this would not be heard by many adult television viewers, the interruption of the line rate at field rate gives rise to 'inter-carrier buzz.'

Incidental phase modulation can be measured by direct observation of the vision signal using an f.m. deviation meter, but few such instruments have sufficient a.m. suppression to handle the video modulation. A more suitable method is to use the television demodulator audio output, provided it does employ the inter-carrier system; this will normally have very good a.m. suppression.

The incidental phase modulation is measured by connecting the demodulator audio output to a noise and distortion analyser, calibrating this with a 50 kHz deviation audio tone, and then noting the spurious deviation produced by various conditions of video modulation of the vision transmitter.

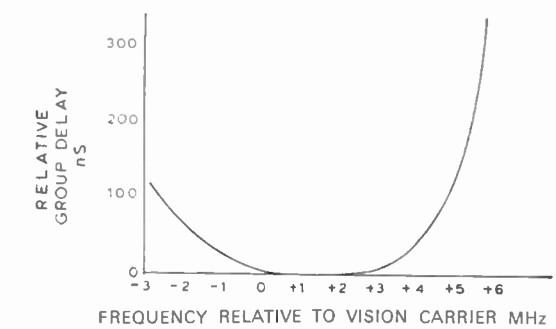
Most specifications require the spurious deviation to be at least 46 dB below the 50 kHz reference deviation for any normal condition of video modulation. The demodulator is allowed the normal de-emphasis characteristic of the audio channel, thus incidental phase modulation at frequencies above the normal audio range is, in effect, ignored. Figure 8 shows the test set-up for incidental phase modulation measurement.

3.5 *Measurement of Mains Frequency Noise*

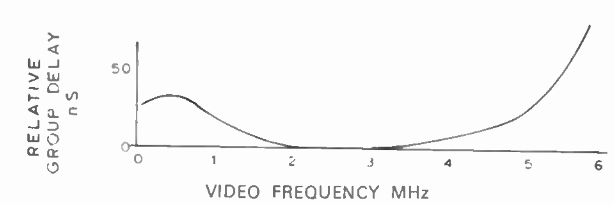
The measurement of mains frequency noise on the output of a high power transmitter can be particularly difficult. Hum loops when measuring noise are well-known to engineers and when measuring adjacent to a three-phase rectifier producing say 20 kV at 6 A to a klystron amplifier, the problem can be acute. A reliable method of measurement is to use a simple diode detector coupled to the output feeder, ensuring that the associated cables are well away from the transmitter mains supply cables. The detector video output is displayed on a battery-operated oscilloscope.

3.6 *Group Delay*

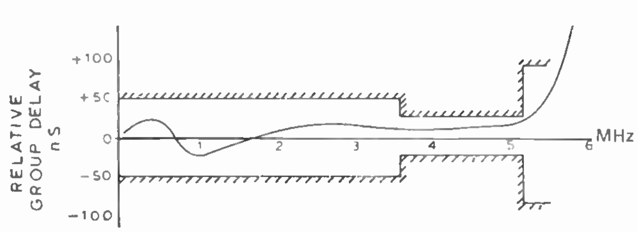
The method of measuring group delay in video systems is generally to use a swept signal technique. The test equipment generates a swept video signal which is amplitude modulated. The delay is measured by comparing the envelope of the swept video signal after passing through the test item with the original signal envelope.



(a) Typical transmitter group delay distortion before correction.



(b) Typical demodulator group delay distortion.



(c) Typical transmitter group delay response, showing specification limits.

Fig. 9.

With measurements on transmitters, however, the introduction of a demodulator again gives rise to uncertainty. The specified demodulator group delay is usually a very considerable proportion of that allowed by the transmitter specification, and hence it is desirable to be able to check it in order that the necessary adjustment to measured results can be made.

This is done using a very broad double-sideband test transmitter sufficiently basic that it may be considered to be incapable of introducing group delay distortion itself.

As with the parameters discussed in 3.3, the distortion produced by a system is distortion in gross form. This is as a result of frequency response shaping at the upper and lower limits of the frequency response to produce the vestigial sideband characteristic, and as a result of the 'notch' at sound frequency produced by the vision-sound combining network. (See Fig. 9.)

4. Other Measurements

In addition to the measurements already described there are a number of parameters which have to be checked on a complete transmitter site which are not specifically related to television, but which are mentioned here because they serve to indicate the vast complex of test equipment required by transmitter commissioning engineers and transmitter operating staff.

These additional measurements include—

Sound checks such as frequency response, harmonic distortion, a.m. and f.m. noise.

Spurious and harmonic radiations.

Carrier frequency stabilities.

5. Special Measurements associated with Parallel or Multiplex Installations

Few transmitting sites have a single television transmitter to carry the service. The most simple arrangement is a main transmitter with a standby transmitter, perhaps with a lower output power, but it is more usual these days to have either a pair of similar transmitters operated in parallel with automatic phasing equipment, or to use a 'multiplex' system.

The multiplex system comprises a single vision and a single sound transmitter incorporating automatic switching arrangements which enable either vision or sound klystron output amplifier to operate as a combined vision and sound amplifier at reduced power output as a reserve service.

Both paralleled and multiplex systems involve additional measurements and more test equipment. The paralleled equipment requires additional measurements of the relative phase of the two separate transmitter outputs. The availability of equipment for direct measurement of relative phase at u.h.f. simplifies this problem. A vector voltmeter can be coupled to the two output feeders using cables of the same electrical length. The meter can be zeroed by initially connecting both cables to a pair of directional couplers at the same point on one feeder, or by using a tee adaptor connected to the one coupler.

The built-in phase detectors of the automatic phasing equipment can be set up in the same way and the transmitter phase indicator can be calibrated against the output of two probes separated by a measured distance on the one feeder, but the vector voltmeter gives an accurate indication at virtually any phase angle.

A multiplex transmitter installation provides a unique measurement problem. As previously mentioned, the 'reserve' operating condition involves one of the klystron amplifiers carrying both the vision and the sound transmission. The inherent non-linearity of a klystron amplifier gives rise to intermodulation products and, in order that these spurious signals are sufficiently small in amplitude, the synchronizing pulse power level under reserve conditions is limited to about one-fifth of the normal power level. The most significant intermodulation product, as far as picture interference is concerned, is a signal which is vision carrier frequency plus sound carrier frequency minus the sub-carrier sideband frequency, a spurious frequency which produces a bar pattern on a picture.

The intermodulation products must be between 50 and 60 dB below the synchronizing pulse level for a satisfactory service.

A 'three tone test' is used to check the level of intermodulation products. Three r.f. signals are applied to the klystron amplifier input, one at vision carrier frequency, one at sound carrier frequency and a third at the sub-carrier sideband frequency.

The levels of these signals are adjusted such that at the output the vision carrier signal is 8 dB below the reserve synchronizing pulse level, the sound carrier signal is 7 dB below the same reference, and the sub-carrier signal is 17 dB below the reference. These levels represent the worst case colour signal which can exist as far as the production of intermodulation products is concerned.

The intermodulation products are readily measured using a spectrum analyser coupled to the output feeder by a directional coupler.

The test signal can be produced by a number of means, but the vision and sound carrier signals are available from the driver transmitter and can be set to the appropriate levels once modulation has been removed. It is possible to derive the sub-carrier sideband signal from a power signal generator which can be conveniently coupled in to the klystron input circuits by using one of the couplers provided at this point for signal monitoring purposes. An alternative method if the vision transmitter can be sine wave modulated is to connect a video oscillator set to sub-carrier frequency to the transmitter video input socket, then by suitable setting of the levels the three tones can be produced from the driver transmitter combined signal output socket.

6. Stability of Line-up

There is an increasing tendency to operate television transmitters by remote control, the transmitters themselves being housed in a vandal-proof building completely unattended.

It is therefore important for all the various parts of the complete installation to function for long periods of time maintaining the specified performance.

When it becomes necessary to prove this capability, as for instance as part of an acceptance test, any of the measurements described in the previous paragraphs may require to be made at random intervals over a period of weeks.

This type of proving becomes as much a test of the measuring equipment as a test of the transmitters. It is essential that all test equipment be capable of accurate measurements a short time after switching on, or alternatively, must be left switched on all the time.

It will be apparent that the set of test equipment required for exhaustive checking of a complete television transmitting site represents a large capital outlay, and it is unusual for a transmitter organization to have all the necessary test gear at every one of their sites. It is therefore vital that any piece of test equipment be able to be transported from site to site maintaining the accuracy required in its use.

7. Acknowledgments

The author would like to thank the Managing Director, Marconi Communication Systems Ltd., for permission to publish this paper.

8. References and Bibliography

1. Wills, R. G., 'U.h.f. aerial and feeder measurements', Proceedings of Conference on Television Measuring Techniques, pp. 405–16, 1970 (I.E.R.E. Conference Proceedings No. 18).
2. Barnes, J. E., 'A test set for measuring television parameters', *Marconi Instrumentation*, **12**, No. 7, pp. 144–8, September 1970.
3. White, N. W., Robinson, K. W. and Heinzl, J. J., 'Measurement of non-linear distortion in television systems', *Marconi Instrumentation*, **12**, No. 6, pp. 23–8, June 1970.
4. White, N. W., 'Monochrome and colour television system measurement', *Marconi Instrumentation*, **11**, No. 3, pp. 16–20, December 1967.

Manuscript first received by the Institution on 7th March 1970, and in final form on 30th April 1971. (Paper No. 1394/COMM 47.)

U.D.C. 621.38/9–523.8

Applying the Phase-locked Loop in Communications and Instrumentation

By

JOHN F. ROULSTON, B.Sc.†

Based on a paper presented at a meeting of the Northern Ireland Section on 9th March 1970 and awarded the Eric Megaw Memorial Prize.

The mode of operation of the phase-locked loop is briefly described, an analytical note being given in an Appendix. Communication applications are the tracking receiver, f.m. detection, the phase-locked homodyne and f.s.k. detection. Instrumentation applications are phase modulation, pulse-width modulation, phase-locked servo for magnetic recording and a phase-locked frequency standard. The voltage controlled oscillator, phase detector and loop filter necessary in these applications are discussed.

1. Introduction

The concept of the phase-controlled oscillator has existed since the 1930s and the subject has been progressively formalized with the growth of control engineering. Early exploitation of the concept was prevented by lack of a suitable mathematical framework for analysis and by the limitations of the electronic devices and the circuit techniques available to early experimenters. Recent interest has been stimulated mainly by the American space programme which brought with it the necessity for a communications system capable of operating over very long range on low transmitter power. To maintain a usable signal/noise ratio at the power levels possible in practice (100 mW for early systems), requires an information bandwidth of about 3 Hz. However at the chosen frequency (around 100 MHz) Doppler shift imposes an uncertainty of some 6 kHz so that a conventional receiver incurs a noise penalty of the order of 2000 times greater. A receiver incorporating a phase-locked loop can lock to the incoming signal over the range of carrier frequencies involved and thus maintain an effective noise bandwidth which is almost equal to the information bandwidth.

Over the past few years many analyses of phase-locked systems have appeared in the literature, so that at the present time there exists a considerable fund of theoretical background to the subject. Unfortunately, however, the reporting of methods of implementing phase-locked systems, the circuitry involved, and the practical application of systems, has been much neglected. As a consequence circuit designers encountering a phase-locked loop for the first time may well experience unfamiliarity with its components and systems designers may well overlook the simplifications that this system can often effect. It is the purpose of this paper to discuss various means of implementing the loop components using modern semiconductor devices and also to point out some less orthodox applications of phase-locked loops. A minimum of theoretical material is presented since the references quoted provide an adequate if not complete guide to the enquiring reader. However for the sake of completeness within the paper a simple analysis of the basic loop is given in Appendix 1.

† Formerly at the Queen's University of Belfast; now with Ferranti Limited, Edinburgh.

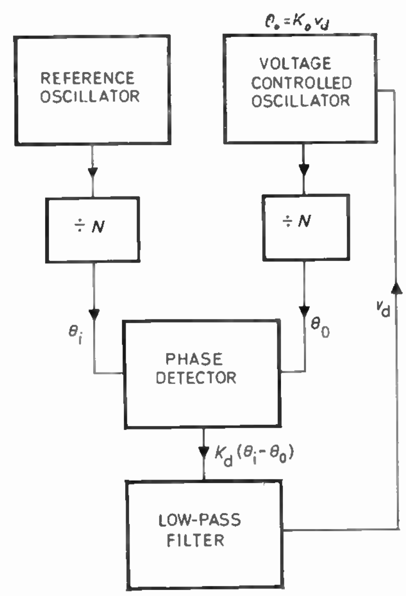


Fig. 1. Basic arrangement of loop components. Counters may be necessary to ensure equal frequencies are present at phase detector inputs.

2. Operation of the Phase-locked Loop

The analytical properties of the general phase-locked loop have been dealt with elsewhere,[1] so that only a brief description of the operation of the system is required here. Figure 1 shows the basic arrangement of loop components. The operation of the loop can be easily understood if initially the assumption is made that this loop is locked, i.e. the phase detector is receiving inputs from the reference oscillator and the voltage-controlled oscillator, (v.c.o.) which are equal in frequency. Under these conditions the phase detector produces an output voltage proportional to the phase difference of its inputs. This output is filtered to remove transients and fed back to control the frequency of the v.c.o. Hence the loop acts as a proportional-controller with respect to phase and in the locked condition the phase error $(\theta_i - \theta_o)$ will be proportional to the reciprocal of open-loop gain. However, since phase angle is the time integral of frequency the loop may also be considered as an integral-controller with respect to frequency, and as is characteristic of this form

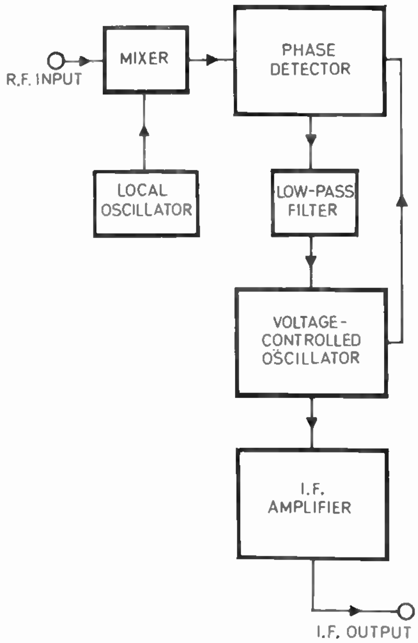


Fig. 2. Scheme for tracking filter receiver.

of control no average frequency error exists in the locked condition. An introductory analysis which outlines the above points is presented in Appendix 1.

In general signals appearing in the loop may be either sinusoidal or digital in character or in some instances it may be advantageous to employ both types of signal. A full analysis of the digital loop with a linear phase detector is due to Goldstein.[2] The sinusoidal phase detector makes the system less mathematically tractable in terms of the prediction of out-of-lock behaviour. A useful analysis has been made by McAleer.[3]

3. Communications Applications of the System

3.1 *Tracking Receiver*

The earliest and perhaps the best known application of the phase-controlled oscillator is its use as a tracking filter in which role it has the capability of recovering signals which are deeply embedded in wide-band noise. The loop acts as a low-pass filter to injected noise. To take best advantage of the system very low information bandwidths are commonly employed allowing loop bandwidths of less than 1 Hz. The loop bandwidth is controlled by the open-loop gain and the filter time-constants, and in very low bandwidth applications it is common to open out the loop response to facilitate initial lock-on by designing the loop filter to be electronically tunable.[3] For this purpose the non-linear characteristics of junction diodes and field-effect transistors are often used. A block diagram of a basic tracking receiver is shown in Fig. 2.

3.2 *F.M. Detection and Threshold Extension*

If the reference source in a phase-locked loop is an f.m. signal and the deviation is such that the voltage controlled oscillator operates over a substantially linear region of its voltage frequency characteristic, then the output of the phase detector is a faithful reproduction of the modulating signal. The operation can be understood by reference to Fig. 3. Assuming that the input signal is of the form,

$$\theta_i = \int (\omega_C + \Delta\omega \cos \omega_M t)\, dt$$

and the loop has a high gain such that $\theta_i \simeq \theta_o$, then ω_0, the frequency output of the v.c.o., is given by

$$\omega_0 = \omega_C + \Delta\omega \cos \omega_M t \qquad \text{......(1)}$$

However, if the v.c.o. has a linear transfer characteristic, i.e.

$$\omega_0 = K_0 v_d \qquad \text{......(2)}$$

then obviously,

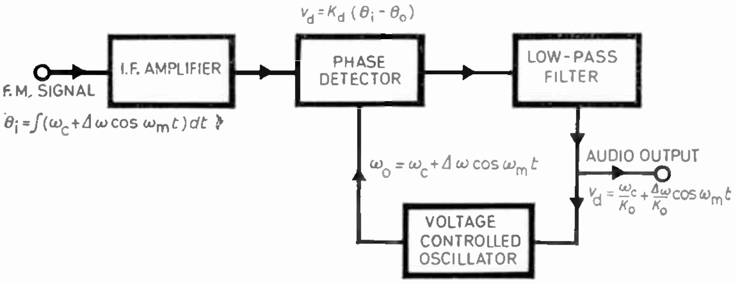
$$v_d = \frac{\omega_C}{K_0} + \frac{\Delta\omega}{K_0} \cos \omega_0 t \qquad \text{......(3)}$$

which represents the modulating audio frequency plus a d.c. level. A full analysis of this system predicts a 3 dB extension of the normal f.m. noise threshold and in practice commercial units have attained signal to noise ratios better than 65 dB with a full range audio bandwidth when using this technique. It is clear that the phase-locked f.m. receiver offers significant performance advantages over its more conventional counterpart and with low cost and ready availability of suitable electronic devices it is becoming commercially attractive.

3.3 *The Phase-locked Homodyne*

The homodyne principle was a topic for experimentation in the developing days of radio before the superheterodyne receiver in its present form became firmly established. Essentially a local oscillator is tuned to the signal carrier frequency and is mixed with the amplitude-modulated carrier. The resulting i.f. is 0 Hz, so that the detected a.m. appears at the mixer output. Unfortunately it is essential for correct operation that the local oscillator should also be in phase with the carrier. Some homodyne receivers achieved this by injecting the r.f. carrier into the oscillator so that it

Fig. 3. Principle of f.m. demodulation by means of a phase-controlled oscillator.



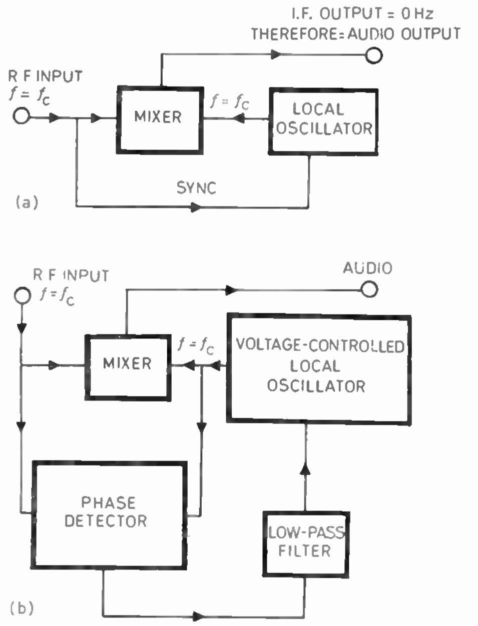


Fig. 4. (a) Original homodyne principle.
(b) Modification to ensure phase coherence.

was effectively pulled into phase. However the direct synchronization method was inherently susceptible to impulsive interference and as a result was never exploited for a.m. detection on a commercial scale. A more practical approach to implementing the homodyne involves using a voltage-controlled local oscillator and including it in a phase-locked loop. The original homodyne is shown in Fig. 4(a), and the phase-locked modification in Fig. 4(b). The system is now immune to interference due to the action of the loop filter which acts as a short-term memory. In physical terms the locked-oscillator acts as a flywheel and maintains its frequency for a short time even in the absence of a carrier signal. At the time of writing it is possible to implement a phase-locked homodyne at very moderate cost using a single special purpose integrated circuit.

This technique is useful in other carrier-recovery situations as for example in the case of coherent detection of single side-band a.m. where a product demodulator requiring a carrier reference signal is used. Similar to this is the recovery of the pilot tone transmitted in multiplexed stereophonic f.m.[4]

3.4 *Detection of Frequency-shift-keyed Data*

The growth of the computer industry has brought with it the need to transfer data in binary form along land lines. In general such lines have a sharp cut-off around 3 kHz so that pulse transmission is impractical over long distances. A convenient method of sending data along telephone lines is to encode the data in the form of a discrete f.m. signal, i.e. bursts of two frequencies, each one being assigned a binary digit. This is termed frequency shift keying. Communications theory predicts that the best method of detecting such signals is by the use of a matched-filter, i.e. a filter whose impulse response is the time reversal of the transmitted signal. A perfect matched-filter (also minimum distance or correlation detector) is not realizable in practice, but it can be approximated by various methods, one of which is a phase-locked loop. The basic system is shown for normal f.m. detection in Fig. 3, and the binary data is extracted at the phase-detector output. For best performance in this application the loop constants are critical and the system is best suited to cases where the keyed frequencies are close together and the data rate is slow. However in such instances it affords a very simple method of decoding.

4. Instrumentation Applications of the System

Integrated circuits, and in particular integrated logic circuits, make possible inexpensive and compact construction of digital loops. The reduction of discrete components is an advantage which can be well exploited in instrumentation applications.

4.1 *Phase Modulation*

A simple modification to the basic phase-locked loop, as shown in Fig. 5, permits its use as a phase modulator. The gain of the loop acts to maintain the output of the differential amplifier at a constant voltage as determined by the frequency of the reference oscillator and the transfer characteristic of the v.c.o. Thus applying a modulating voltage change Δv_m to one input of the differential amplifier causes a change $-\Delta v_m$ to occur at the other input, which in turn implies a phase shift of the v.c.o. signal with respect to the reference oscillator of $-\Delta v_m/K_d$. The linearity of the system is governed only by the linearity of the phase detector and the differential amplifier and can be made excellent if a digital phase detector with a sawtooth characteristic is used. The modulating range possible is determined by the range of the phase detector and may be increased by a factor of N by including divide-by-N counters in the loop as

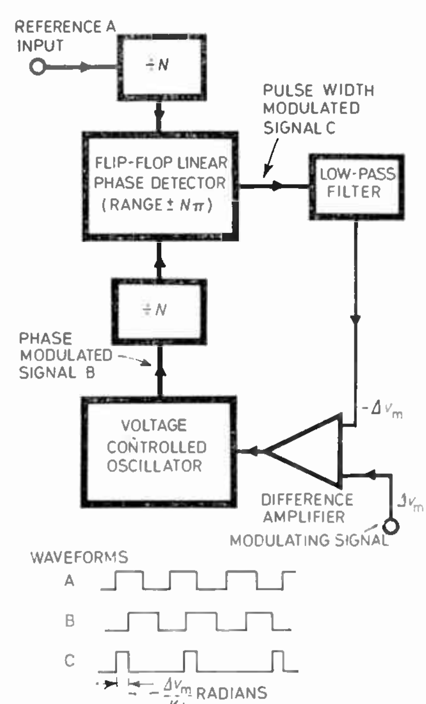


Fig. 5. Use of loop for phase modulation and pulse-width modulation with waveforms.

indicated in Fig. 5. The technique affords phase modulation over a very wide range.

The author has successfully used the phase modulator as a basic element in the analogue simulation of the radiation characteristics of phased linear aerial arrays. Treating the aerial as an array of point sources the far-field radiation pattern can be derived in the form,

$$R_p = \sum_{i=0}^{n} A_i \exp\left\{j\left(\omega t + \phi_i + \frac{2\pi d_i \sin\theta}{\lambda}\right)\right\}$$

which is a summation of phase modulated terms. An n-element array is simulated by $(n-1)$ loops driven from the same master oscillator. In the author's simulator a completely digital approach was employed and the square-wave outputs of the v.c.o.s filtered to produce sinusoids, before summing to produce the field vector, R_p.

4.2 *Pulse-width Modulation*

The arrangement of Fig. 5 is also capable of very accurate pulse-width modulation. If the filter is made sufficiently wide to give the loop a level response over the audio range, and an audio frequency signal is applied at the input of the differential amplifier, phase modulation of the v.c.o. with respect to the reference source occurs as before. However, if the phase detector is a set reset flip-flop designed to trigger on, say, the negative edges of its input waveforms, the mark/space ratio of the phase detector output is defined by the phase difference of the reference and v.c.o. waveforms and hence is very closely proportional to the modulating signal. The author has used this method of pulse-width modulation over a dynamic range up to 90% of the reference source period. Linearity is difficult to measure experimentally over such a wide range but a figure of 0·1% seems a conservative estimate. The pulse-width modulator has been used as a servo amplifier by driving switching transistors from the phase detector waveform. Very high powers are available in this mode (typically 20 W from 200 mW transistors), since the output devices are either in cut-off or saturation. Plans are underway at the time of writing to evaluate the technique for high-power audio amplification.

4.3 *Phase-locked Servo for Magnetic Recording*

The techniques discussed in the last Section lend themselves well to the design of a phase-locked servo-system of the type used in digital magnetic recording. Fig. 6 outlines the basic idea. Tape speed is selected by choosing the reference frequency. The need for a servo amplifier to drive the d.c. servo can be eliminated by using the pulse-width modulation properties of the phase detector output. The tachometer may consist of a photo-electric or magnetic sensor and filtering in the loop is provided by the mechanical inertia of the capstan and flywheel.

4.4 *Phase-locked Frequency Standard*

The B.B.C. Droitwich transmission from Rugby is a frequency standard transmission, with a stability of the order of one part in a thousand million. At a carrier frequency of 200 kHz on the long wave band this transmission makes an ideal reference source to which a digital oscillator may be phase-locked to produce a laboratory frequency standard. Such an instrument has been constructed using a loop bandwidth of 0·5 Hz. It is interesting to note that this loop will not lock to a 200 kHz signal from a common commercial signal generator because the instability of the generator exceeds the loop bandwidth.

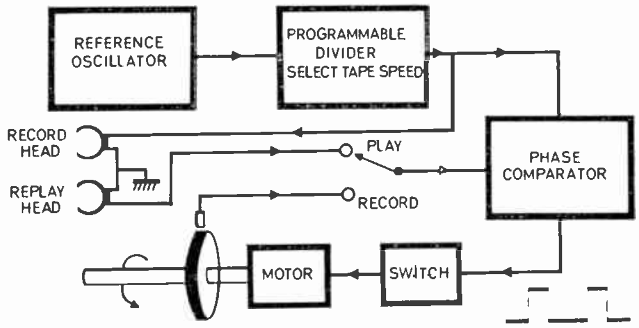


Fig. 6. Phase-locked servo loop for magnetic recording in aircraft systems and using pulse-width modulation to control servo-motor.

In the design of such an instrument it would be unwise to contemplate a purely digital approach since this would necessitate limiting outside the loop, say immediately after a high-Q 200 kHz carrier filter. Limiting outside the loop in this case degrades the signal/noise performance attainable and would reduce the stability of the final output. Instead, a switching type phase detector which can cope with one sinusoidal input and one square wave input should be chosen. A suitable circuit is described in Section 5.2.

5. Methods of Implementing Loop Components

5.1 *The Voltage-controlled Oscillator*

In practice there are many ways of implementing the v.c.o. A list of the methods tried by the author are given below:

(1) Collector-coupled astable with variable voltage on timing resistors.

(2) Collector-coupled astable with controllable current sources charging timing capacitors.

(3) Emitter-coupled astable with variable current sources in emitters.

(4) Unijunction transistor with controllable current source charging timing capacitor.

(5) Cross-coupled TTL monostables, modulate voltage to which timing resistors are connected.

(6) Varactor tuned L-C oscillator.

(7) Phase-shift oscillator using f.e.t.s as voltage-variable resistances in ladder network, Wein bridge or twin-T.

(8) Phase-shift oscillator using photo-resistors with filament lamps in epoxy block as variable resistances.

(9) Phase-shift oscillator using non-linear dynamic impedance of junction diode as voltage variable resistance.

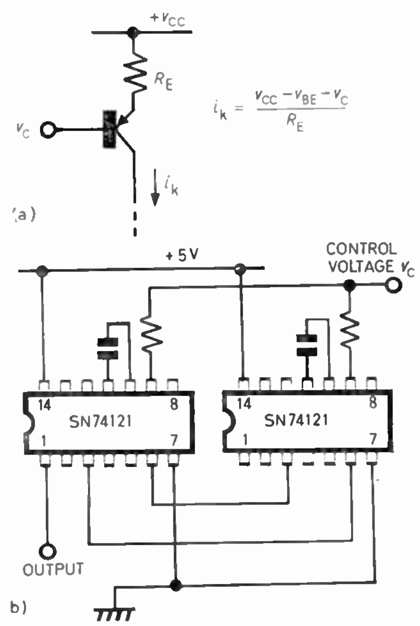


Fig. 7. (a) Current source for controlling astables.
(b) Integrated circuit v.c.o. from standard TTL packages.

The circuit of a current source suitable for use in cases (1) to (4) is given in Fig. 7(a) and Fig. 7(b) shows the cross-coupled monostable arrangement. The important oscillator constant as far as the loop is concerned is the effective gain expressed in radians per second per volt and given the symbol, K_0. This is estimated by the gradient of the frequency voltage curve evaluated at the working frequency.

5.2 *The Phase Detector*

For phase detection at r.f. or i.f. the balanced diode phase detector[6] is useful. However this form of detector has the disadvantage of requiring two transformers and at the lower frequencies these could become bulky components. A very simple method of phase detection which has been used by the author at frequencies up to 1 MHz is shown in Fig. 8. The circuit is basically a shunt f.e.t. chopper and is very suitable for situations where the reference signal is sinusoidal and the v.c.o. is digital. An interesting feature is the possibility of locking to harmonics of the v.c.o. range, and the author has locked a digital v.c.o. centered at 200 kHz to a 1 MHz reference sinusoid using this circuit. The circuit when operated with inputs of the same frequency has a range of $\pm 90^\circ$ and obeys a cosine law. Operating with multiple frequencies increases the range but correspondingly decreases the sensitivity.

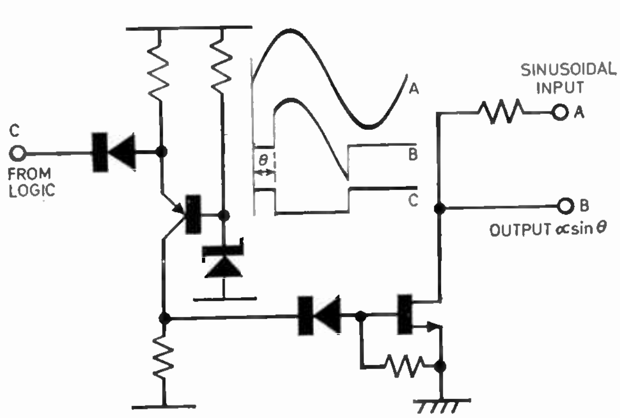


Fig. 8. Chopping phase detector suitable when only one signal is in digital form.

The simplest of all phase detection processes occurs in digital systems where the detector may be a pulse-triggered set reset flip-flop. This form of phase detector has a linear or sawtooth characteristic which is repeated with a periodicity of 180°. Extending the phase detector range is accomplished by frequency division of its inputs, a division of N-times in input frequencies resulting in an extension of N-times in the operational phase detector range. The phase detector constant, K_d, is the gradient of the transfer characteristic measured over the linear operating range and has the dimensions of radians/volt.

5.3 *The Loop Filter*

The choice of filter determines the order of the loop. A first-order filter implies a second-order loop which has guaranteed stability. Third- and higher order loops are uncommon and are difficult to stabilize. Thus, in general, a lead-lag filter, as shown in Fig. 9(a), is used. This arrangement has the advantage that its two time-constants τ_1 and τ_2 permit the loop natural radiancy, ω_n, and the damping factor, ζ, to be chosen independently. This filter can also be used in its active form as in Fig. 9(b), and the possibility of gain is an advantage.

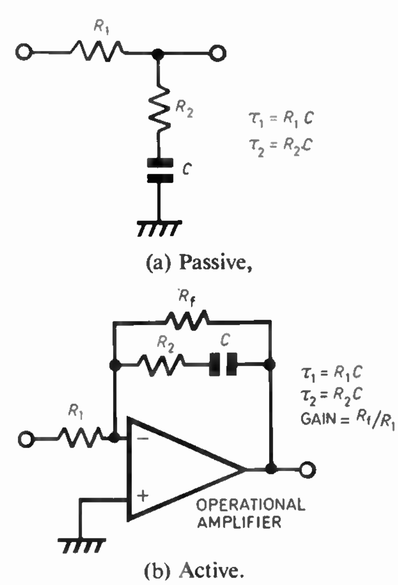


Fig. 9. Loop filters

6. Conclusion

The phase-locked oscillator as a system has been shown in a wide variety of applications and implemented in a comparably wide variety of ways. Care has been taken to cite only low-cost systems such as may have a ready outlet in commercial equipment. The technique has even wider applications in more specialized areas as, for example, the phase-locking or mode-locking of lasers and the phase-locking of high-power microwave oscillators such as cavity tuned magnetrons to high-stability low-power references. The author feels that a wider appreciation of the scope of this technique could lead to system simplifications and reduced equipment cost in many instances.

7. Acknowledgments

The bulk of the study forming the background of this paper was undertaken when the author was an undergraduate student in the Electrical Engineering Department of the Queen's University of Belfast. The author expresses sincere thanks to Dr. J. McBride and Dr. A. Refsum who jointly provided the stimulus for the study and provided many hours of valuable discussion, and also to his colleague Mr. D. Donaghy for much practical assistance.

8. References

1. Gardner, F. M., 'Phaselock Techniques'. (Wiley, New York, 1966.)
2. Goldstein, J., 'Analysis of the phase-controlled loop with a sawtooth comparator', Bell System Monograph 1023.
3. McAleer, H. T., 'A new look at the phase-locked oscillator', *Proc. Inst. Radio Engineers*, **7**, No. 6, pp. 1137–43, June 1959.
4. Portus, R. T. and Haywood, A. J., 'Phase-locked stereo decoder', *Wireless World*, **76**, No. 1419, pp. 418–22, September 1970.
5. Perry, M. A., 'Magnetic recording; latest developments in ground and airborne systems', *Electronic Engineering*, **42**, pp. 48–51, September 1970.
6. Krishan, S., 'Diode phase detectors', *Electronic and Radio Engineer*, **36**, No. 2, pp. 45–50, February 1959.
7. Woodvine, R., 'The phase-locked loop', *Microelectronics*, pp. 13–18, July 1970.
8. Camenzind, H. R. and Grebene, A. B., 'Phase-locking: integrated tuned circuits made easy', *Electronics*, **45**, pp. 94–8, 28th April 1969.
9. Gill, W. L., 'Use i.c.s in your phase-locked loop', *Electronic Design*, pp. 76–80, 11th April 1968.
10. Leonhardt, R. and Fleischmann, H. H., 'Pull-in range of phase-lock circuits with arbitrary feedback filter', *The Radio and Electronic Engineer*, **36**, No. 2, pp. 101–10, August 1968.
11. Rey, T. J., 'Automatic phase control: theory and design', *Proc. Inst. Radio Engineers*, **48**, No. 10, pp. 1760–71, October 1960.

9. Appendix

Considering Fig. 1, the phase detector output voltage is proportional to the phase difference of its inputs,

$$v_d = K_d(\theta_i - \theta_o) \qquad \text{......(1)}$$

K_d is called the phase detector gain factor. Phase error voltage is filtered by the low-pass filter which also suppresses noise and high-frequency signal content. Filter transfer function is $F(s)$. The v.c.o. has the transfer function,

$$\omega_0 = \dot{\theta}_o = K_0 v_2 \qquad \text{......(2)}$$

By taking Laplace transforms equation (2) becomes

$$\theta_o(s) = \frac{K_0 V_2(s)}{s}, \qquad \text{......(3)}$$

which says that the v.c.o. output phase is proportional to the integral of the input control voltage. Also,

$$V_2(s) = F(s)V_d(s), \qquad \text{......(4)}$$

and transforming (1),

$$V_d(s) = K_d(\theta_i(s) - \theta_o(s)) \qquad \text{......(5)}$$

These combine to give the basic loop transfer relationship,

$$H(s) = \frac{K_0 K_d F(s)}{s + K_0 K_d F(s)} \qquad \text{......(6)}$$

Assuming $F(s)$ is passive lead-lag filter such that

$$F(s) = \frac{s\tau_2 + 1}{s(\tau_1 + \tau_2) + 1}, \qquad \text{......(7)}$$

then $H(s)$ becomes

$$H(s) = \frac{s\omega_n(2\zeta - \omega_n/K_0 K_d) + \omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \qquad \text{......(8)}$$

where the natural radian frequency, ω_n, and damping factor, ζ, are given by the expressions:

$$\omega_n = \sqrt{\left(\frac{K_0 K_d}{\tau_1 + \tau_2}\right)}, \qquad \zeta = \frac{1}{2}\sqrt{\left(\frac{K_0 K_d}{\tau_1 + \tau_2}\right)} \cdot \left(\tau_2 + \frac{1}{K_0 K_d}\right)$$

Manuscript first received by the Institution on 4th November 1970 and in final form on 22nd March 1971. (Paper No. 1395/CC104.)

U.D.C. 621.396.62

H.F. Communication Receiver Performance Requirements and Realization

By

B. M. SOSIN,
B.Sc., C.Eng., F.I.E.E.†

Presented at a meeting of the Communications Group held in London on 7th October 1970.

Discusses h.f. communication problems and the stringent and often conflicting requirements placed on the performance of any h.f. communication receiver. The H2900 receiver has been designed with these requirements in mind. The receiver uses a novel method of digital frequency synthesis allowing for small frequency increments of 1 Hz.

1. Introduction

The well-known advantages of the 2 to 29 MHz spectrum for long-range communication has resulted in each of the 9000 channels of 3 kHz bandwidth being fully occupied. For optimum quality on each channel it is essential to use as much of the 3 kHz bandwidth as possible. For optimum reliability in communication receivers it is equally essential that any channel can be selected in the presence of possible interference from other channels and from noise. Furthermore, the receivers should not cause interference with other channels.

The quality of a communication receiver may be defined in terms of the performance achieved under these conditions.

The translation of this simple statement into a specification of the various receiver parameters involves conflicting requirements, which necessitate compromises in design. In spite of modern techniques and design aids, this does lead to difficult decisions regarding the choice of optimum compromise. In the following sections conditions under which a receiver operates will be analysed and the effect of various parameters on the performance indicated. The design of the H2900 receiver will then be presented.

2. Performance Considerations

2.1 *Gain*

The receiver gain is not a problem nowadays and any required gain can be obtained easily.

2.2 *Noise Factor*

The noise factor of the receiver may limit the reception of a number of channels if the signal strength is insufficient. However, some of these channels may already be unusable due to atmospheric, man-made or galactic noise. From the various charts given in the C.C.I.R. Report No. 322 two curves were selected Fig. 1, curve A giving the condition in high-noise area at noisy times and curve B low-noise area at quiet times. Of course, occasionally the incoming atmospheric, man-made and galactic noise will be higher than curve A and lower than curve B. However, these are exceptional circumstances and since a communication system must be designed for a reliable reception, conditions better than those represented by curve B should not be assumed. These curves are translated in Fig. 2 to show the relation between the number of available channels and the channel signal strength using 'ideal' receivers. The limitations imposed by the external conditions are severe.

The C.C.I.R. report gives the value of atmospheric, man-made and galactic noise above the thermal noise. Discussions on receiver performance must be with respect to a level of input signals at the receiver terminals, because this gives much clearer comparisons. For this purpose the power level of the atmospheric, man-made galactic noise can be referred to the thermal noise level in a terminated 50 Ω feeder and expressed in the equivalent signal level above 1 μV. This reference is used in Fig. 2 and in the following discussions.

It is also necessary to define what is the usable signal. This would of course depend on the type of service, i.e. whether it is speech or data, also on the tolerable quality for a particular user, i.e. military or commercial. In the following discussion it has been assumed that a signal level of 0 dB and above will give a signal/noise ratio which is acceptable.

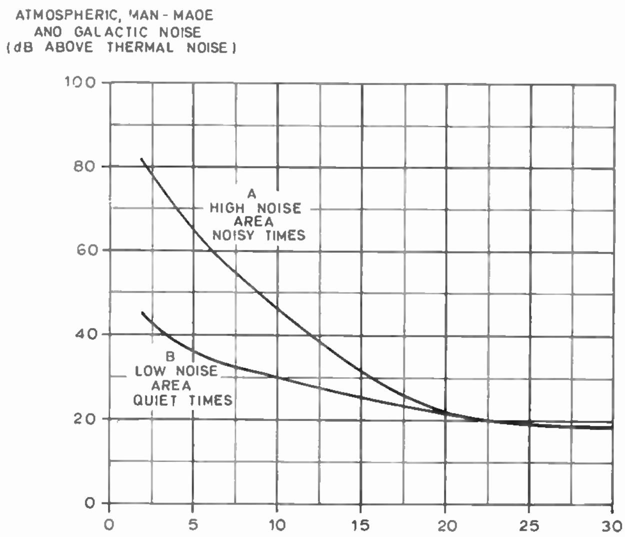


Fig. 1. Expected atmospheric, man-made and galactic noise.

† Marconi Communication Systems Ltd, Chelmsford, Essex.

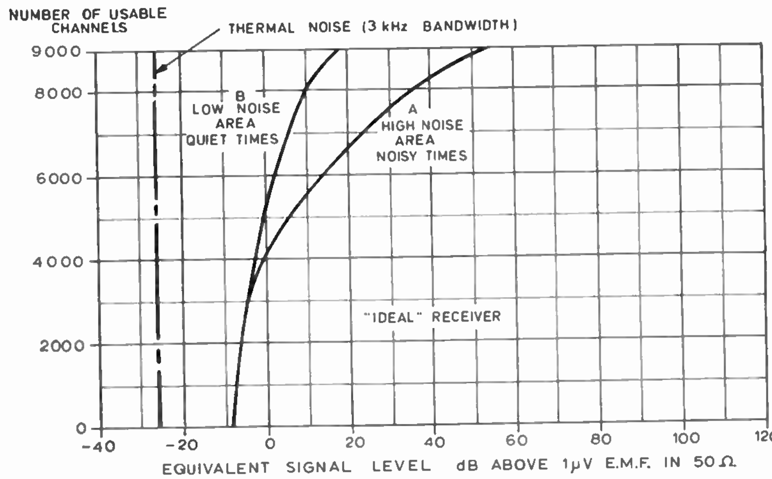


Fig. 2. Effect of expected atmospheric, man-made and galactic noise on number of usable channels. 0 dB signal/noise ratio. 0 dB antenna gain.

A practical receiver will be connected to an antenna, where the gain of the latter, together with atmospheric conditions, will determine tolerable noise factor of the receiver selecting three types as below:

(1) An antenna with no gain but incurring 6 dB distribution loss in feeders.

(2) Antenna with 14 dB gain (rhombic or equivalent) incurring also 6 dB distribution loss. A passive multi-coupler providing 8 outlets is used, increasing distribution loss by another 9 dB to a total of 15 dB.

(3) A 16 ft whip antenna where a 5 dB mismatch loss at 30 MHz and 10 dB mismatch at 7 MHz is assumed.

Figure 3 shows the allowed receiver noise factor when connected to the above antennae. It is evident that a 10 dB noise factor is better than needed. In the past there has been quite a lot of controversy as to whether a receiver should have a noise factor of 7 dB or even 5 dB. In the opinion of the author these arguments are irrelevant. Present technology allows the design of a receiver with these very low noise factors, but this is neither needed nor desirable. The noise factor, the reciprocal mixing and intermodulation products (i.p.s) are parameters mutually conflicting. Improvements in noise factor usually make the i.p.s worse and it will be shown later that this is not desirable.

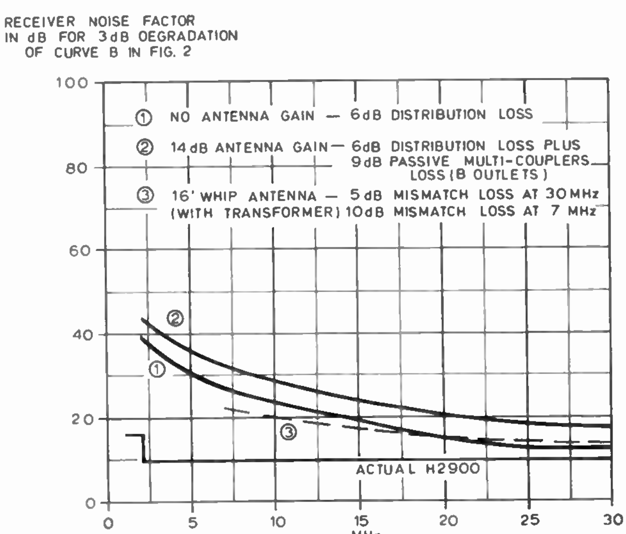


Fig. 3. Allowed noise factor.

2.3 *Interference Signals*

A multiplicity of high level interfering signals will be present at the input terminal of a receiver due to overcrowding of the h.f. band. These will vary from location

Table 1. Assumed input signals

Bands of frequencies MHz ±500 kHz	Number of assumed signals in each band at the levels of: 40–50 dB (μV)	50–60 dB (μV)	60–70 dB (μV)	70–80 dB (μV)	80–90 dB (μV)	90–100 dB (μV)
2·5	16	8	4	2	1	1
3·5	16	8	4	2	1	1
4·5	24	12	6	3	2	1
5·5	30	16	8	4	2	2
6·5	30	20	10	5	3	3
7·5	30	20	10	5	2	3
8·5	30	20	10	5	2	3
9·5	30	20	10	5	2	3
10·5	30	20	10	5	2	3
11·5	30	20	10	5	2	2
12·5	30	20	10	5	2	2
13·5	30	20	10	5	3	2
14·5	30	20	10	5	3	2
15·5	30	16	8	4	2	1
16·5	30	16	8	4	2	1
17·5	24	12	6	3	1	1
18·5	16	8	4	2	1	1
19·5	16	8	4	2	1	0
20·5	16	8	4	2	1	0
21·5	16	8	4	2	1	0
22·5	8	4	2	1	1	0
23·5	8	4	2	1	1	0
24·5	8	4	2	1	0	0
25·5	8	4	2	1	0	0
26·5	4	2	1	0	0	0
27·5	4	2	1	0	0	0
28·5	2	1	0	0	0	0
Total	546	321	160	79	38	38

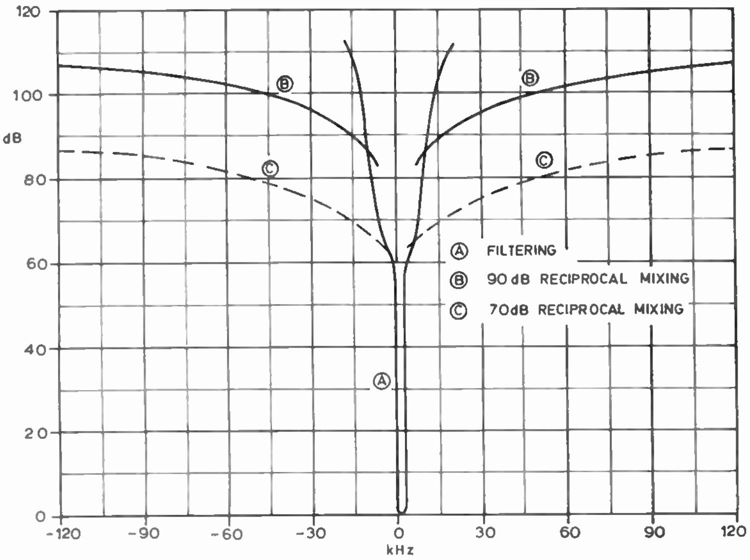


Fig. 4. Filtering and equivalent filtering due to reciprocal mixing.

to location, season of the year and time of day. In order to establish the effect of these interfering signals on the useful channels one must first assume a likely condition under which a receiver will be working. For this purpose a number of reports have been studied as well as measurements made at various sites.

Table 1 gives the likely number of signals at different strengths in each 1 MHz band between 2 and 29 MHz. This Table will now be assumed as standard in considering parameters such as filtering, reciprocal mixing and intermodulation products.

2.4 *Filtering*

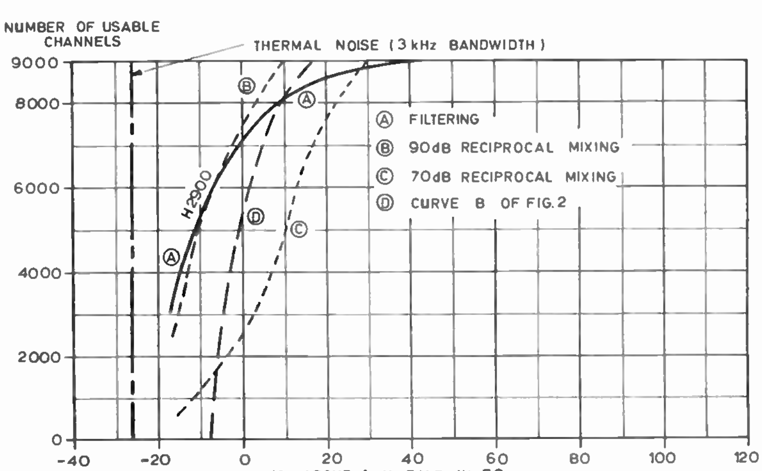
Curve A on Fig. 4 shows a receiver with a very good filtering characteristic obtained by a sideband filter and a so-called roofing filter. Most, but not all, high-quality communication receivers would be expected to have a similar performance. In spite of this excellent filtering a number of channels will be unusable due to interference. Curve A in Fig. 5 gives the number of usable channels as limited by the filtering. Curve D is re-plotted from Fig. 2 for comparison with the effect of the external noise.

2.5 *Reciprocal Mixing*

Only relatively recently has a phenomena called 'reciprocal mixing' been appreciated. In superheterodyne receivers the radio frequency is converted to intermediate frequencies in various mixers. Under normal conditions incoming signals are mixed with locally-generated heterodyne signals. However, no signal is noise free. When a large interfering signal appears at the mixer it will also mix with the noise of the heterodyne signal and, although the interfering signal may be out of the i.f. band, the noise so produced will be inband. This phenomena is called reciprocal mixing and is measured by the amount of noise introduced by a closely spaced interfering signal, i.e. when the level of the interfering signal, expressed in dB above 1 μV, spaced 20 kHz away from the wanted signal will produce a reduction of signal/noise ratio by 10 dB.

One can consider reciprocal mixing as an equivalent path by-passing i.f. filters and this is shown on Fig. 4 for 90 dB and 70 dB reciprocal mixing. The number of usable channels for these cases it is shown on Fig. 5. It is clear that a receiver with high reciprocal mixing will

Fig. 5. Effect of insufficient filtering and reciprocal mixing on number of usable channels (0 dB signal/noise ratio).



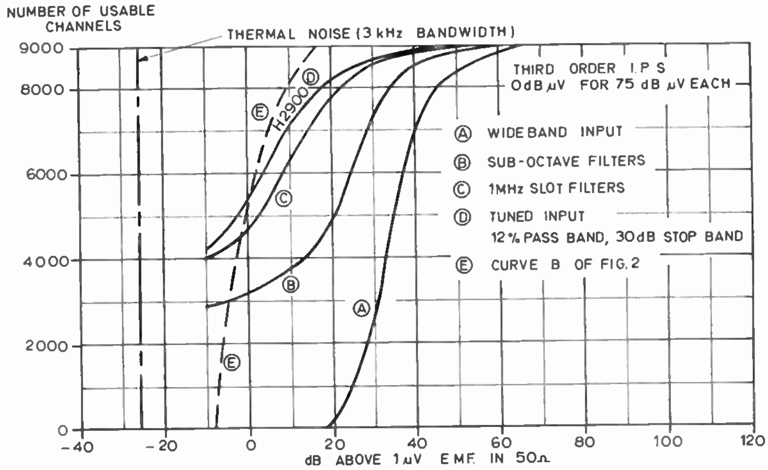


Fig. 6. Effect of 3rd-order intermodulation products on a number of usable channels (0 dB signal/noise ratio).

perform well. However, 70 dB reciprocal mixing will introduce severe limitations on the receiver performance. It is of interest to note that only very few receivers have this parameter better than 70 dB and many considerably worse. Such receivers will appear very noisy indeed under operating conditions, and an operator will confuse this with the high noise factor of a receiver, which it is not.

2.6 *Intermodulation Products*

When a number of interfering signals are present in the input to the receiver they will mix with each other and produce intermodulation products (i.p.s). This is due to non-linearities in the amplifying or frequency conversion devices. This phenomena can have a profound effect on the performance of a receiver. The most important i.p.s are those produced by the third-order non-linearity.

Consider first a receiver with excellent linearity defined by a case where two interfering signals of 75 dB (μV) e.m.f. produce third-order i.p.s of 0 dB (μV) e.m.f., which varies in proportion to a level of signals of at least 110 dB (μV) e.m.f. Assume the same number of input signals at various levels as when discussing the reciprocal mixing, i.e. those given in Table 1. A number of i.p.s can now be calculated and related to the number of available communication channels as a function of the signal strength of these channels. The results are plotted on Fig. 6 for various types of input circuits to the receiver. A completely wideband circuit produces intolerable situations (curve A) where almost all channels are blocked at the input level of as high as 20 dB (μV) and even with a signal of 30 dB (μV) a high proportion of the channels are blocked.

The situation improves when sub-octave filters are used (curve B) but even here approximately half of the available channels are blocked when incoming signals are below 20 dB (μV). Slot filters of 1 MHz bandwidth produce a considerably improved situation (curve C). To obtain a really good performance the receiver input must be tuned. However, the input circuit will not have a perfect stop-band rejection and in the calculation it was assumed that the stop-band attenuation is only 30 dB, whilst the bandwidth of the passband is 12%. Curve D expresses the result, which is quite good.

It is evident that even with this excellent linearity an input tuned circuit is required. Many receivers do not have such a good linearity and to make them operate satisfactorily under present-day conditions considerably better input tuned circuits are required. In order to be able to use sub-octave filters the linearity of the receiver would have to be increased to 100 dB (μV) (see Fig. 7). Departing from a receiver consideration, active antenna

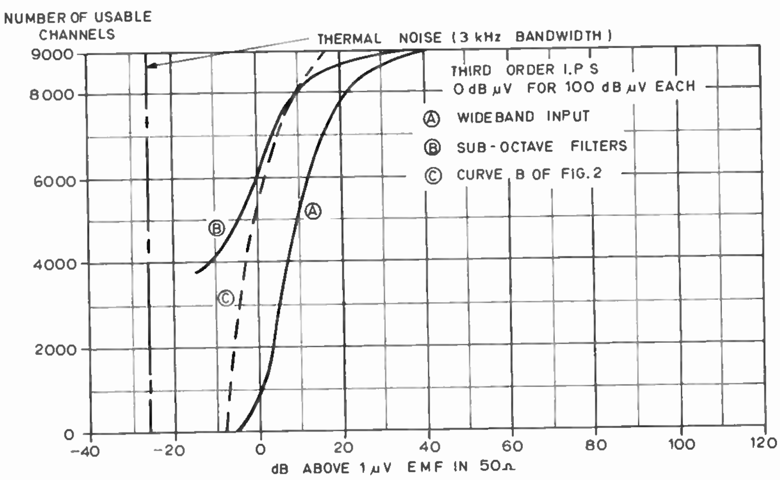


Fig. 7. Effect of 3rd-order intermodulation products on a number of usable channels (0 dB signal/noise ratio).

multi-couplers must be wideband. In order that they do not seriously effect receiving systems their third-order i.p.s must be less than 0 dB (μV) for the signals of at least 110 dB (μV).

The second-order i.p.s are of course very much less in number and they are not present at all when sub-octave or MHz slot filters are used. With 30 dB stop-band attenuation of the input tuned circuits they will have but negligible effect on the receiver performance using a tuned input. They will however affect receivers with wideband inputs and in considering multi-couplers they must not be forgotten.

2.7 *Frequency Stability*

High stability and accuracy in frequency setting are essential for two main reasons. In the first place with the overcrowded h.f. spectrum it facilitates quick location of the required communication channel. This becomes even more important in the case of remote operation of the receiving stations, or in the case of short messages in ground-to-air or military applications.

Many modern communication systems, e.g. data, v.f.t., Piccolo, Lincompex, etc., require a very high order of frequency accuracy, 1 Hz or better, and a frequency synthesizer is essential.

2.8 *Other Effects*

Two other parameters often discussed are crosstalk and blocking. Both of these parameters are related to the i.p.s and reciprocal mixing. In the opinion of the author if the latter two are satisfactory, crosstalk and blocking will be satisfactory automatically.

During various frequency conversions in the receiver there are bound to be some spurious signals. By proper choice of the various i.f.s and good screening they can be minimized. The heterodyne signals as derived from the frequency synthesizer must also be free from spurious frequencies. The permissible level is exceptionally low and in general they should be below the level of those from reciprocal mixing.

3. H2900 Solid State Receiver†

The above discussion indicates clearly the target performance of a high-quality communication receiver. The H2900 series of receivers were designed with the object of this target. The fully equipped receiver can be described as:

- dual diversity
- triple i.f. superheterodyne
- synchronous detector
- independent sideband and
- digitally-controlled heterodyne oscillators.

The receiver has solid-state active devices throughout. By extensive use of microcircuits and miniature components, it has been possible to assemble the receiver into a container of 17·8 × 44·6 × 56 cm (7·0 × 17·5 × 22 in). Compare this with past receivers which occupy a 2·13 m (7 ft) cabinet.

The block diagrams (Figs. 8 and 10) show the basic arrangement and more detailed descriptions and justification of the design are given below.

Figure 8 gives a block diagram of the signal path where, following the input circuits, the frequency is converted to the first i.f. of 79·8 to 81·8 MHz. The second i.f. of 30 MHz is used and contains a crystal roofing filter. After the final conversion to 2 MHz, the signal is split to upper sideband, lower sideband and carrier paths, amplified, detected and followed by audio amplifiers. Separate a.g.c. is used in each sideband except where carrier a.g.c. is employed. Only a small amount of a.g.c. signal is applied to the earlier stages of the receiver.

Figure 9 shows the signal levels through the receiver and it clearly indicates that very little amplification is provided in the early stages of the receiver. Before final filtering of the sideband, the signal level is kept as low as possible without appreciably increasing the noise factor of the receiver.

† See also Sosin, B. M., 'A breakthrough in h.f. receiver design', *Point-to-Point Communication*, 4, No. 1, January 1970.

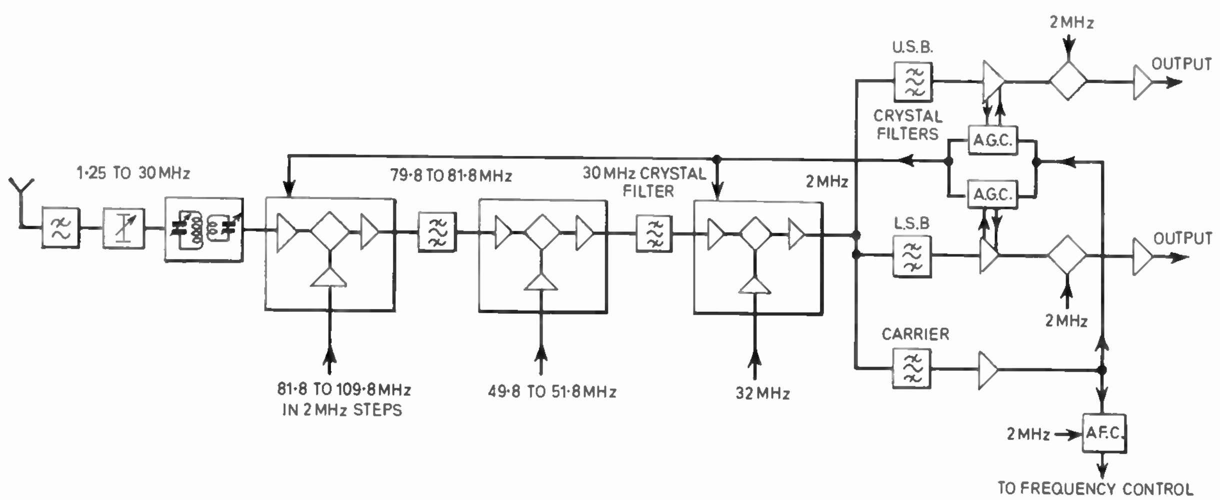


Fig. 8. Block diagram—signal path.

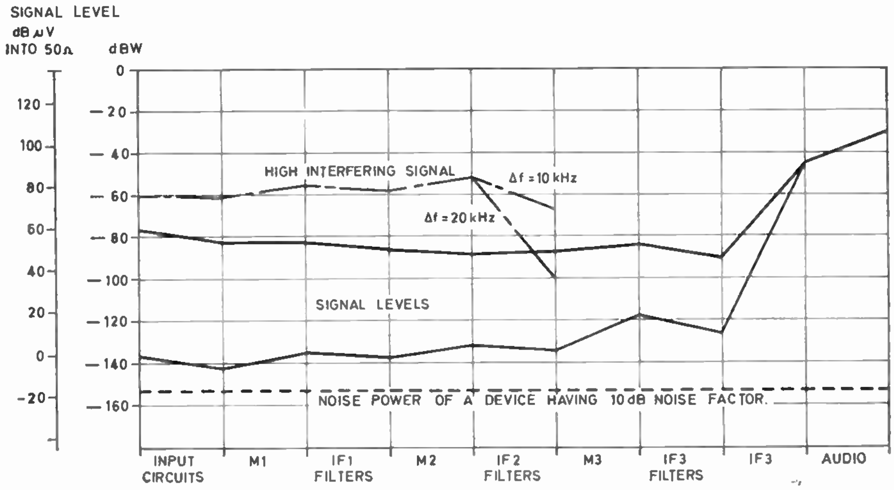


Fig. 9.
Signal level through the receiver.

The synthesizer block diagram on Fig. 10 shows the method of deriving the various heterodyne signals, i.e.:

81·8 to 109·8 MHz in 2 MHz steps,

49·8 to 51·8 MHz continuously variable,

32 MHz and

2 MHz.

All of these frequencies are referred to the external standard frequency. A detailed description follows in Section 3.5.

3.1 *Choice of Intermediate Frequencies*

The choice of intermediate frequencies needs justification where one must consider the advantages and disadvantages of single, double and triple superheterodyne. Advances in the design of crystal filters allows the construction of sideband filters working at 2 MHz. Choosing the final intermediate frequency (IF_3) of 2 MHz eliminates the necessity of additional conversion to the region of 100 kHz.

In a receiver with a single i.f. the radio frequency is converted directly to a relatively low intermediate frequency. Even if the latter is as high as 2 MHz, which will inhibit reception round about 2 MHz, in order to avoid image reception the requirement on the input filtering is impossible. This scheme must be rejected outright.

The next, less simple arrangement would be to use a high first i.f. above the highest receiver frequency, followed by a relatively low second i.f. The situation is considerably improved in this case but closer examination indicates that the amount of isolation between the first and second frequency changer is again almost impossible to obtain in practice. Lack of this isolation will cause image receptions at twice the second i.f. Reception of two input signals spaced at the second i f. will also be present. This arrangement also requires the heterodyne oscillator to be continuously variable over the 30 MHz band from a frequency synthesizer. Although it is possible to combine the two high-frequency signals generated in the synthesizer (see Section 3.5) into one continuously variable signal, the mixing process involved would be detrimental to the reciprocal-mixing performance of the receiver. This is another reason for avoiding a dual frequency-conversion system.

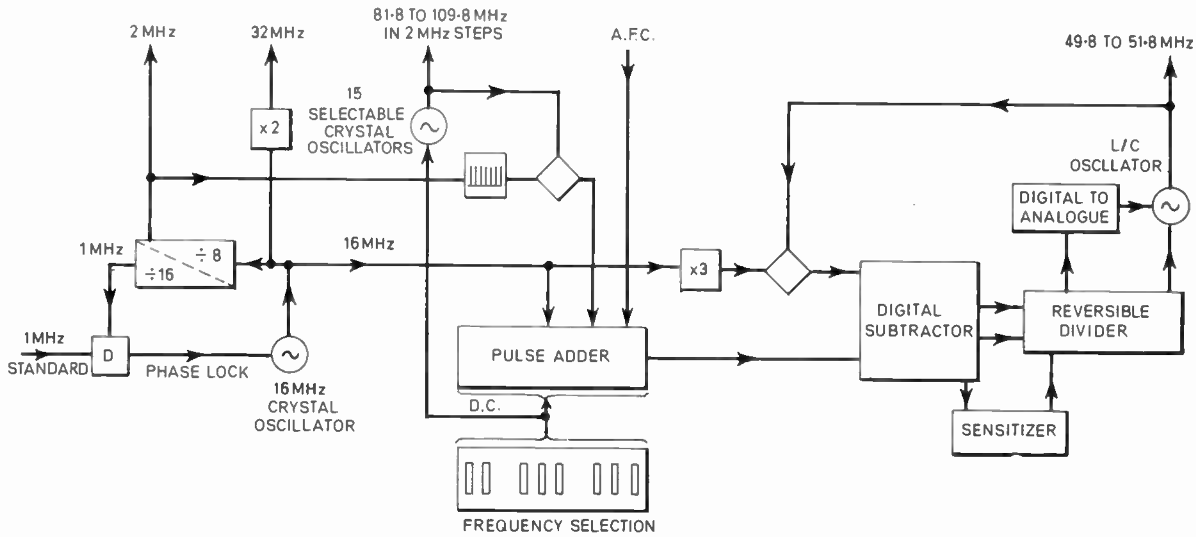


Fig. 10. Block diagram—frequency control.

Triple-frequency conversion is thus necessary. Restraints on the i.f. and heterodyne frequencies can now be placed in the usual manner taking into account harmonics, particularly harmonics of the heterodyne signals to the high orders. The analysis led to the five possible selections (Table 2).

Table 2

Choice of intermediate frequencies

	f_1	IF_1	f_2	IF_2	f_3	IF_3
1.	84 to 112	82 to 84	68 to 70	14	16	2
2.	90 to 118	88 to 90	74 to 76	14	16	2
3.	70 to 98	68 to 70	42 to 44	26	24	2
4.	82 to 110	80 to 82	50 to 52	30	32	2
5.	84 to 112	82 to 84	52 to 54	30	32	2

The selected schemes were then investigated in greater detail. I.p.s up to the 15th order were plotted and weighted according to their order and fractional separation relative to the intermediate frequencies. They were also checked by a computer. No particular scheme is outstandingly advantageous but generally of the similar pairs 2 is preferred to 1, and 4 to 5 as the lowest-order spurious frequencies are further off-tune. On the remaining possibilities 2, 3, 4, scheme 3 is an inverting mode which has some slight disadvantages for a drive application and 2 requires the highest frequency for the variable oscillator. Scheme 4 is therefore used as the best compromise. This scheme also has the possible advantage that all the intermediate and oscillator frequencies employed are either outside or near the extreme limits of the h.f. band.

3.2 *The R.F. Tuner*

The r.f. tuner is a bandpass filter with a double resonant circuit which is automatically tuned, working between nominally-matched impedances.

The total frequency range is divided into five octaves. Frequency setting is by means of digital switches on the front panel of the synthesizer, selecting the required range via high-reliability sealed relays. After the selection of the range by certain digits, the remaining digits define the position of the variable capacitor. However, before the digits can be used they must be changed into digits representing directly the capacitor position, i.e. they must be code converted. This conversion is performed by digital microcircuits and the actual position of the capacitor obtained from a shaft encoder. If the required and actual positions are not the same, a motor brings the two into alignment.

In view of the number of possible non-mechanical systems available, the choice of a variable capacitor may appear surprising. Before making this decision the following alternative possibilities were investigated:

(1) sub-octave filters,

(2) slot filters,

(3) permeability tuning, and

(4) varactor tuning.

Sub-octave and slot filters are clearly unsuitable, as discussed earlier.

Tuning by varying inductor permeability with high magnetizing currents has also been tried, but suffers from hysteresis effects, temperature changes and what is more important, introduces non-linear distortion.

By using large-capacitance varactor diodes it was possible to design a good workable system but the performance was limited by non-linearities being introduced. As the complication and relative cost of this method were higher than those of simple mechanical tuning, the latter was adopted.

3.3 *Frequency Converters*

The design of frequency converters for linear operation requires great care. It is a subject on its own. The assembly consists of a broadband f.e.t. amplifier working in a grounded gate connexion, followed by a diode bridge mixer using Schottky (hot-carrier) diodes, followed by another f.e.t. amplifier. The heterodyne signal is amplified to a fairly high level before application to the bridge mixer.

3.4 *Automatic Gain Control*

An a.g.c. system of sophisticated design is incorporated to ensure successful operation in bad fading conditions. Each sideband has a separate circuit so that it is possible to receive signals satisfactorily with as much as 30 dB difference between the incoming sideband levels. Two possible systems are available.

3.4.1 Carrier/sideband system

With this system the carrier is extracted by means of a crystal filter and used as the level-controlling signal for both sidebands. The principal application is for speech reception but it can be used also for data. However, it is generally preferable to derive level-controlling information from the sideband during reception of data. For this reason a sideband a.g.c. is also incorporated with relatively fast attack time-constant (but slower than is customary in order to avoid the receiver blocking on the bursts of noise) and a slow decay time-constant. With these time-constants chosen for optimum data or speech reception, a sudden burst of signal or noise would overload the final stages of the receiver and possibly cause trouble on a telephone circuit connected to the output. To overcome this, an additional very fast acting a.g.c. sub-system is also connected, to function as a limiter. By reducing the decay time of this third system to a minimum a burst of interference does not block the receiver.

3.4.2 The sideband system

Many operators, particularly in a military environment, do not use a carrier in transmission and a carrier a.g.c. system is then of no value. However, with the long decay time necessary for speech, sideband a.g.c. does not allow a quick re-establishment of communication with a weaker station after a conversation with a strong station has been concluded (this requirement is necessary for 'on-net' working). For these circumstances an a.g.c.

inhibiting circuit has been included which operates a few seconds after the cessation of the speech signal. Data and fast override a.g.c. operate in a manner similar to that in the carrier/sideband systems.

3.5 *Synthesizer*

The 16 MHz internal crystal oscillator is locked to the external 1·0 MHz standard and is used to control all the frequencies within the receiver. Both the 32 MHz and 2·0 MHz frequencies are derived from the 16 MHz oscillator (Fig. 10).

A total variation of some 30 MHz in the heterodyne oscillators is necessary to bring any r.f. signal to a fixed i.f. A study of various oscillators reveals that a mechanical or voltage-variable oscillator covering such a wide frequency range would not be sufficiently stable, nor microphony-free to be used in a high grade receiver, particularly for data reception. Therefore, this receiver incorporates a variable L-C oscillator covering a range of only 2·0 MHz, the complete r.f. range being covered by crystal oscillators in 2·0 MHz steps.

The crystal oscillators are not phase-locked to the standard, but are off-set by 200 kHz from integral frequencies. This offset (a deliberate error) plus the actual error in frequency is separated by comparison with multiples of 2·0 MHz standard frequencies, and is applied as a correction to the variable L-C oscillator, via the digital control circuits.

Neither the selectable crystal oscillators nor the L-C oscillators are placed in an oven as might be expected. Although an oven could be constructed to keep the temperature within a few millidegrees Celsius, this would require a long time to stabilize itself after switching on. Furthermore, the oscillators would have to work at a temperature above the highest likely ambient temperature, thereby giving a higher noise factor and lower reliability. Because of the close frequency control, long-term stability of these oscillators is not necessary. The oscillators are mounted in containers having a very long thermal time-constant, the actual time-constant of the L-C oscillator being six hours. In order to keep the switch-on stabilizing time short, the power dissipation within the oscillator is very small, so that the change of temperature, due to internal heating, produces only negligible drift after 15 minutes (see Fig. 11).

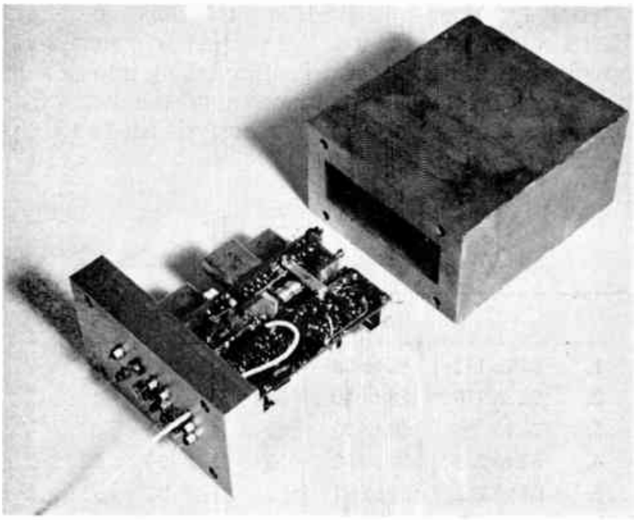


Fig. 11. The precision L-C oscillator with its solid container.

Digitally-controlled synthesizers commonly employ a variable-divider principle, in which the frequencies of the standard and the controlled oscillators are divided down to the frequency of the highest common denominator, i.e. the lowest frequency step. An entirely different principle is used in the H2900 receivers. In the first place a pulse train with a number of pulses per second (corresponding to the chosen frequency) is derived by digital techniques using 16 MHz as a clock frequency. Careful attention to logic sequences provides pulses at an almost uniform rate, but in spite of this an effective phase movement of some few hundred degrees is present. The frequency error of the crystal oscillator, as well as a.f.c. correction if used, is now added. Note that a.f.c. corrects the setting of the synthesized frequency and does not use free running oscillators. The pulse train so derived corresponds to the required frequency of the L-C oscillator, when the latter is brought back to the range of 1·8 MHz to 3·8 MHz by subtracting 48 MHz.

A pulse subtractor is used for comparison. It is important to add sufficient backlash in this subtractor

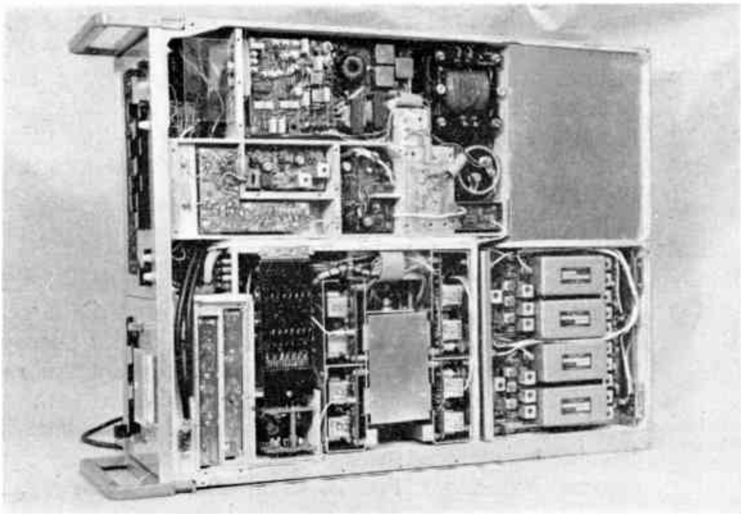


Fig. 12. Receiver with top cover removed.

to allow for phase movement in the assembled pulses as indicated above. The frequency error is now integrated in a reversible binary divider, converted to an analogue signal in a digital-to-analogue converter and applied to control the L-C oscillator. The servo loop so constructed is equivalent to a phase-locked loop but without reversal of control voltage at every 360° phase point. Such a loop will always pull into frequency. It is of particular interest to note that, because of the presence of backlash in the subtractor, under normal conditions the servo loop is open and closes only for brief intervals whilst the frequency is being corrected.

The sensitizer is a refinement of the system to reduce the time required to pull the L-C oscillator into the correct frequency after a change.

3.6 *Remote Control*

A complete control of all receiver functions is permissible from an extended or remote position. The remote panel may be almost identical to the local panel. Such a control requires some 70 bits of information to be transmitted. However, by means of information storage combined with conversion into a serial form, only seven wires are required for extended control. All the information storage and code conversions are performed by digital microcircuits.

The distance for the remote operation is not limited, neither is the transmission medium. By means of a suitable data-transmission equipment interposed into the seven wires, the receiver may be operated via telephone, radio or even a satellite link.

3.7 *Construction*

The receiver with top cover removed (Fig. 12) shows easy accessibility to modules in spite of the compact assembly. The remaining modules are accessible from the other side. In order to improve reliability plugs and sockets have been eliminated to the absolute minimum. The layout of the front panel is shown on Fig. 13.

Fig. 13. Front panel layout.

4. Acknowledgments

The author wishes to thank the Technical Director of Marconi Communication Systems Ltd. for permission to publish this paper. The design of a receiver of this complexity is the result of work of a team of engineers with particular contribution to Messrs. M. H. Hopcraft, E. J. Conley, D. G. Ely and J. H. Cone.

Manuscript received by the Institution on 30th November 1970. (Paper No. 1396/Com 48.)

U.D.C. 535.653: 621.397.62

A Standard Colour Monitor Matching Comparator

By

RAY E. KNIGHT†

Reprinted from the Proceedings of the Conference on Television Measuring Techniques held in London from 12th to 14th May 1970.

This paper surveys the problem of matching colour television monitors and the relative merits of optical or electronic aids. An optical design for assessing the normalizing white of a colour monitor is then discussed at length, and comments are made in detail on the alternative solutions at each stage of the design. It is concluded that a Lummer-Brodhun matching field, together with a tungsten lamp and glass filter combination for the reference D_{65}, form an economical solution to the problem of a standard reference illuminant. Design flexibility allows the instrument to be used for other applications such as checking film review screens, and shadow mask tube phosphors.

1. Introduction

Early in 1968, Thames Television chose to develop an instrument to set the white point of colour monitors accurately to D_{65}. Every attempt had already been made in our studio installation, and in the procedures for lining-up picture sources, to ensure that equal RGB signals originated from neutral test scenes. Work had been carried out on the perceptibility of RGB errors as seen in the reproduction of flesh colour[1] and it was concluded that a broadcaster should be able to set-up the white point of colour monitors just as accurately.

With continued experience, television operators concerned with colour matching, in common with industrial colorists, become sensitive to colour errors that will otherwise go unnoticed by other people. It becomes necessary to offer these specialists the best possible colour matching facilities at certain stages of their work in order that they will happily continue to do a good job, even if this means working inside the tolerances necessary for a good broadcast service.

Development work was therefore carried out between Thames Television Ltd., Michael Cox Electronics Ltd. and Grafikon (Engineers) Ltd. together with the services of the National Physical Laboratory, Teddington. Grafikon have subsequently made production quantities of this instrument based on the prototype. The design aim was to achieve a critical matching field with a *standard* colour reference, at an attractive price.

2. Monitor Layouts

It is current practice to place colour monitors next to monochrome monitors in preview monitoring arrays, the nett effect being to provide an adaptation field near these colour monitors in the order of 10 000 K. If a colour monitor is set to glow at 6500 K, colour adaptation will make it look pink within the high colour temperature surround.

Solutions to this are: (1) use all colour monitors—expensive; (2) set the colour monitors' colour temperature to equal that of the monochrome monitors—wrong by definition[2]; (3) filter the monochrome monitor light—feasible, but this requires that a very low value of ambient light falls on the filter if its colour is not to become visible; (4) use a monochrome c.r.t. with a phosphor that glows at D_{65}—a good solution but not available in Europe when this project was commenced; (5) it is possible to set the colour monitors mid-way between 6500 K and 10 000 K, thereby avoiding some of the unpleasant pinkness, but of course this is still technically wrong.[3]

Other areas such as quality check, camera set-up etc., have isolated colour monitors and these can be set precisely to D_{65} without any of the above disadvantages.

3. Choice between an Optical or Electronic Instrument

Some preliminary work had already been carried out to make an electronic RGB reading tristimulus comparator in an attempt to get an objective and sensitive means of setting the white point that discounted an operator's psychological opinion of a colour match. Initial results were quite encouraging, but it was expedient to set a colour monitor to D_{65} first in order to calibrate the electronic instrument. The basic interpretation of the system specification with regard to white point has to satisfy a standard 1931 CIE observer; this is done most conveniently, visually, with a real observer. The decision was therefore taken to develop a visual instrument as priority and continue development of an electronic model. Operationally, the performance of a visual instrument depends, among other things, on an operator's colour vision and his aptitude at colour matching. This may be a disadvantage. On the other hand, the performance of the eye can be good at detecting small differences of chromaticity and brightness, providing these errors are presented to the operator in a good visual field.

Problems to solve in an electronic aid are stability of colour reference, low light sensitivity and immunity from ambient light, wrong monitor settings from monitors using different phosphors if the instrument does not have standard observer analysis characteristics, and the difficulty in producing the same optical analysis in each instrument on a production run. The magnitude of this latter problem may be judged from the fact that present-generation broadcast-quality colour cameras seldom colour match after a grey line-up, due to economical errors that remain in the optical analysis, and small instabilities in the electronics.

† Engineering Research Group, Thames Television Limited, Teddington, Middlesex.

4. Types of Colour Match

An analysis of colour matching may be useful because the performance of a complete television station goes beyond looking at isolated colour monitors to wondering why this picture here, looks different on that monitor there. Work by Crawford[4] and Wyszecki and Stiles[5] is particularly helpful here.

4.1 *Purely Spatial Match*

If two fields are closely juxtaposed with an imperceptible line of division between them, an operator will be able to match with the minimum of uncertainty and maximum precision, e.g. in the use of a television split screen or circle effect.[4]

4.2 *Spatio-temporal Match*

If two visual fields are separated spatially by an intervening space, large or small, neutral or distracting, a psychological factor comes into play because the two fields have to be remembered, albeit sometimes for a brief interval. Television examples would be comparing two monitors side by side or turning away from a monitor to look at the original scene through a colour-corrected window.[4]

4.3 *Pure Memory Match*

Here, an operator first looks at and becomes statically adapted to one field, then repeats the process for another field, with a sufficient time interval between the two observations to remove all sense of temporal or spatial change. The comparison is by an effort of memory only. The psychological factor is now a maximum—the precision of the comparison a minimum. In television this would be the comparison of a monitor in one control room with that in another after a walk down a corridor; viewing the set scenery from the studio floor and then going into a control room to see its reproduction, or watching a television replay at home having been present at the studio recording session.[4]

4.4 *Symmetrical Match*

If, in a comparison, the result of which is determined by an operator's judgement of what he sees, one scene or stimulus is changed for another, without affecting the result, the two stimuli are said to be visually equivalent by strict substitution. The substitution must be exact; that is, the alternative test stimuli must be imaged on the same retinal area, for the same duration and at the same point in time in the operation, all other stimuli—the conditioning stimuli— must be the same. It follows then that 'A is equivalent to B' (A equiv B) and B equiv A mean exactly the same thing. Similarly, the transitivity law—if A equiv B and B equiv C, then A equiv C—holds strictly.[5]

In television terms the instantaneous change between two pictures on one picture monitor would appear to fall into the above category, because it is possible in practice to match several sources sequentially on a single monitor and these sources will then appear to match in any order on another monitor. Exactly the same thing is expected of the colour monitor comparator. It is necessary for several monitors to match among themselves after each one has been adjusted with the aid of the comparator.

4.5 *Asymmetrical Match*

If, however, the outputs from several sources all looking at a common object, are viewed sequentially on one monitor in order to match their performance, the matching is not strictly symmetrical if the operator's gaze has to follow a changing point of interest from picture to picture, or if different details are included other than the point of interest from picture to picture, etc. The matching definition is now asymmetrical.[5]

4.6 *Short-term Memory Match*

In looking over these analytical definitions with a practical television situation in mind it is useful to note that memory matches are usually completed more quickly than simultaneous ones, with less consistent results than for simultaneous, although they could still be within a satisfactory tolerance limit for colour television adjustments; and that such matches can result in higher values of luminance and saturation in a memory time as short as 5 seconds.[6]

A colour reference placed permanently near a colour monitor will, using a short-term memory match, result in a match being carried out more quickly than an instrument with the best spatial matching field; but, of course, the match will not be as accurate.

5. Types of Matching Field

The view offered to an operator, in which he can see the two stimuli to be compared and matched, is called a matching field. There are many optical ways of bringing together two stimuli, ranging in simplicity and cost, ease of use, and ambiguity of resulting match.

Matching fields are first classified as being either monocular or binocular. In the first case, one field is presented to one eye, usually requiring the observer to look into an eyepiece, which embarrasses people who wear spectacles. The proper use of an eyepiece does define the limits of the matching field, e.g. by excluding extraneous light. A binocular matching field usually implies the presentation of a different field to each eye. Then there is the method of looking at one matching field with both eyes, without eyepieces. While offering little protection from light sources at the edges of vision, it is easy, comfortable and quick for all to use including spectacle wearers.

5.1 *The Bi-partite Field*

This is usually a rectangle or circle divided down the centre. The reference and test stimuli to be compared are seen side by side in each half. An early example of such a field is the Jolly Block. It consists of two blocks of white wax separated by a thin piece of metal. The two sources to be compared each illuminate one of the wax blocks and the observer compares the diffused light either side of the metal strip. Another early type of field is by Conroy and is made by simply placing a 45° reflecting surface in front and to one side of another, the observer looking past the near surface to see the far one.

Both of these examples may be seen at the Science Museum, London, while variations of the latter form are in practical use today in several commercial instruments.

5.2 *The Lummer-Brodhun Photometer Cube*

This is probably the most widely known matching field, and in appearance to the observer consists of a rectangle of reference light with a small ellipse of the test light in the centre. Such a cube is made by putting each hypotenuse of two right angle prisms into contact with each other, having first etched away the hypotenuse of one to leave a raised disk. Where this disk makes optical contact with the other prism, light can pass straight through forming a test spot. The surrounding field is reflected by total internal reflection at the un-etched hypotenuse. A special version of this is called a Lummer-Brodhun contrast head. This is a rectangular bi-partite field, with a low contrast insert of the other side in each half. The matching criterion is to get equal contrast in each half of the field, to which the eye is a very sensitive judge. Unfortunately, it is rather complicated to manufacture.

The colour monitor calibrator uses a Lummer-Brodhun photometer cube, presenting a spatial image to the operator without an eyepiece. This was chosen because it was thought to be slightly more sensitive than a simple bi-partite field, while the reference rectangle can be relatively large in order to act as a partial adaptation field. Another advantage is that the eyes' gaze can wander round the test spot thereby avoiding retinal fatigue that can otherwise occur in a bi-partite field due to the permanent exposure of the reference and test patches on the same part of the retina during the matching procedure.

6. Sources of Reference Illumination

It is conventional when reproducing a nominated colour temperature to aim at a chromaticity as close to the Planckian locus as possible and having a smooth spectral distribution similar to a black body radiator. Tungsten lamps meet this requirement, but are limited to low colour temperatures, unless some form of colour correction is used, because tungsten melts before reaching 3700 K.

Examples of sources that are fundamentally of a high colour temperature are: electroluminescence, xenon arc, ultra-violet radiated paint[7] and fluorescent lamps.

Electroluminescence was not investigated, although a unique construction of a matching field may be possible using a 'panel of light'. Small xenon arc lamps are available but they generate too much light and heat for our instrument, although their colour may be independent of applied voltage.[8] The use of u.v. light to cause a special paint to radiate visually suggests a cover radiating D_{65} over the monitor screen with a cut-out through which to see the monitor light; however, this was not investigated. Fluorescent lamps are now manufactured for colour matching purposes, meeting the requirements of D_{65} within a tolerance acceptable for colour appraisal within the paint, textile and other industries. However, a fluorescent tube produces light from both a mercury arc discharge and the tube's fluorescent coating. The colour and brightness will vary considerably during its initial warm up and slowly throughout its life. These are due to changes of the mercury vapour operating pressure within the lamp which affects the mercury discharge; bulb temperature changes that alter the amount of the phosphor fluorescence; and an increasing mercury deposit inside the tube, forming a pale yellow filter that mainly affects the mercury discharge light as the tube ages. These variables were considered too many for an instrument we wish to call *standard* that would be in daily use. In addition to the above disadvantages peculiar to fluorescent tubes, there are two general disadvantages with the sources mentioned in this paragraph: (1) their colour is nominally fixed and cannot be easily corrected if it is lowered by an efficient diffuser that will scatter blue light, and (2) their spectral distributions are not smooth.

7. Choice of Miniature Tungsten Halogen Lamp

After deciding to use a tungsten lamp, the natural thing was to search the manufacturers' catalogues for small types of lamp. Although some appear satisactory as far as dimensions and light output go, their long life is attributed to a filament temperature below Illuminant A. They will, of course, reach Illuminant A but with both a shortened life and increased blackening of the glass envelope. A search through tungsten halogen lamps then revealed a miniature low-wattage sample. Tests indicated that this particular lamp could be under-rated to produce Illuminant A without any adverse effects while having a useful life in the order of 150 hours. The gas in this lamp contains bromine, which does not selectively absorb the filament light in the visible region of the spectrum as does iodine, therefore the spectral distribution is close to a black body radiator. In actual fact, the chromaticity of a tungsten lamp lies on the green side of the Planckian locus, as does the chromaticity of D_{65}.[9] Iodine gas could make the lamp glow pink.

8. Types of Colour Correction

In order to raise the tungsten colour temperature to 6500 K a spectrally selective filter is required. It is possible to use a laboratory-type Davis Gibson liquid filter, but these are not stable with time and are messy for operational applications, as they are liable to leak. Gelatine filters may be used but their long-term batch-to-batch tolerance is not good enough for a standard and they do not react favourably in the presence of heat. A spectrum template could be used to form the wanted spectral distribution precisely but this would no doubt cause the instrument to become too large to be hand held. Finally, glass filters offer a good all-round solution within the restrictions of glass homogeneity and glass manufacturing tolerances. Special glasses are available that very closely reproduce a black body spectral distribution, but they are foreign to the U.K. and are too expensive.[10] Chance 0B8 glass was therefore chosen because it is freely available in the U.K., and will produce a correlated colour temperature of 6500 K free from sharp spectral bumps.

9. Design of the Glass Filter

The calculation of the filter thickness required to raise Illuminant A, 2855·6 K, to the correlated colour temperature of Illuminant D_{65}, 6504 K, was carried out by Dr F. J. J. Clarke at the National Physical Laboratory, Teddington. It had been decided to operate the lamp at Illuminant A because that is a standard colour which also turns out to be a good compromise between lamp life and the amount of heat generated inside the instrument. The overall procedure was to measure the spectral transmission of a sample of 0B8 glass that was known to be too thick, calculate and predict the correct thickness, reduce the sample to this thickness, and re-test.

In detail, the spectral data was processed in a special digital computer program developed by the N.P.L. Metrology Centre's Colour and Photometry Section which predicted the optimum thickness for this particular sample in order to produce the closest colour match to D_{65}. This was done by evaluating the internal spectral density of the glass, using the spectral variation of refractive index of a typical crown glass to allow for reflexion losses. The effects of trial variations of thickness were evaluated by application of Bouguer's Law, i.e. assuming that the glass was homogeneous with respect to uniformity of distribution of colouring material. At each trial the chromaticity co-ordinates (u.v.) were evaluated in the 1960 CIE Nearly Uniform Chromaticity Scale chart, and the distance from the target co-ordinates of D_{65} ($u = 0{\cdot}19783$, $v = 0{\cdot}31222$) were tested. A Fibonacci Search algorithm was used to establish the minimum possible distance in the (u.v.) chart, and hence optimize the filter thickness.

This sample was reduced to the first predicted thickness and re-tested as above, when it was found that the applied correction was not quite enough. This was thought to be due to stratification in the glass in the sense that the surface layers were not so optically dense as the inner ones: such stratification had also been noted by earlier N.P.L. workers. However, with the mired shift of two thicknesses of a single sample available, it was possible to predict a second thickness, and have another filter ground to this thickness (from a reserved batch of glass) and carry out the test procedure again, as outlined above, in order finally to determine the thickness of production filters. One of these production filters has now been standardized at N.P.L. and is about a just-noticeable difference of chromaticity off target.

10. Production of Even Reference Field and Control of Brightness

The tungsten lamp has to provide a very evenly illuminated reference field, otherwise the presence of shading would worry the observer and reduce the precision of his match.

Reasonable evenness may be achieved by treating the lamp as a point source and situating it far enough away from a diffusing screen. For this to be effective the lamp ends up at such a distance that a very high wattage is required. The brightness of this screen may be controlled by varying the lamp to screen distance, but for a good brightness range the instrument becomes too big, even if a folded light path is used and all the mirrors are made movable.

A mirror box solves the problem of evenness in a short distance[11] and is so good that a mechanical iris may be used to control the brightness of the reference field by vignetting the input face of the mirror box without upsetting its output evenness. Such an arrangement is used in the colour monitor comparator. To increase the brightness range control of the iris, an opaque spot is suspended just in front of the closed iris opening. (The closed iris action is limited to about 2 mm diameter.) In this way, a colourless form of brightness control is introduced, capable of smoothly reducing the reference field to zero brightness.

The mirror box is actually a solid block of clear perspex, with totally diffuse perspex cemented on to each end. Total internal reflexion at the four clear perspex-to-air surfaces reflects the diffused input light forward in kaleidoscope fashion to form the even reference field.

11. Optical and Mechanical Properties of Perspex

The spectral transmission of clear perspex is extremely flat throughout the visible spectrum,[12] but that of totally diffusing perspex is somewhat yellow due to the selective scattering of blue light. However, this may be compensated for by increasing the colour temperature of the source in order to get an equivalent to Illuminant A out. This is an excellent reason for using a tungsten lamp.

The mechanical properties of perspex allow easy machining and polishing methods to adequate standards for this instrument. The possibility of quickly cementing perspex together offers great scope to the mechanical designer.[13] Perspex costs and weighs less than glass for an equivalent complexity of design: the weight saved in this instrument amounts to 0·36 kg (0·8 lb).

Another property of perspex, not taken advantage of in this current design, is the possibility of moulding complex forms to high optical standards.

12. The Basic Instrument Layout

An electro-brightened aluminium mirror is placed behind the lamp, making the lamp system more efficient. The iris restricts the diameter of the cone of light passing to the mirror box and therefore acts as a brightness control. The tungsten light first passes through the A to D correction filter before meeting the input diffuser of the mirror box. Light scattering into the mirror box at the exit of the input diffuser plus the multiple internal reflexions within the box, together with further scattering at the exit diffuser, produces a very even reference field.

A 45° reflecting surface in the photometer block shows the observer the reference field, while a small hole in this surface allows him to see the colour monitor light as a test spot in a good spatial matching field. A partially diffusing screen is placed on the test side of the photometer block to defocus the shadow mask phosphor dots and raster line structure. This diffuser is illuminated by a solid perspex light pipe over a sufficient area to enable binocular vision to see an even test spot at viewing distances greater than 30 cm.

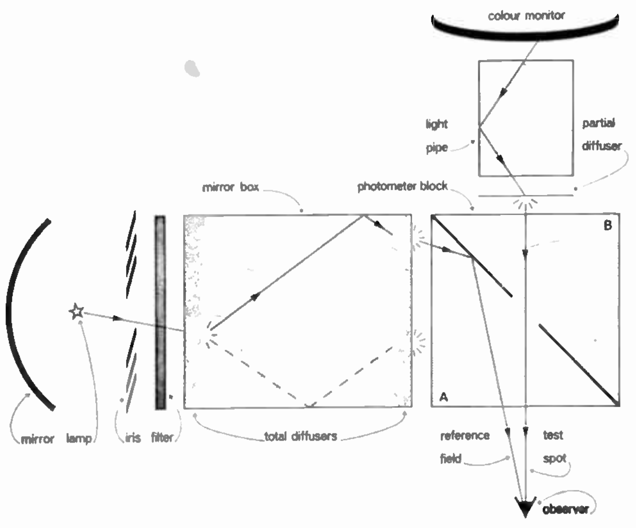


Fig. 1. Schematic diagram illustrating the principles involved in the comparator.

The photometer block is of a unique design. Prism A (Fig. 1) has an aluminized hypotenuse reflecting the reference light to the observer while optical contact is made between the two prisms A and B, at a spot where there is no aluminium, with a little glycerine. A number of concentric channels in prism B serve to prevent migration of the glycerine and allow for its expansion and contraction with changes of temperature.

The photometer block has been made large enough to be easily viewed without an eyepiece, which greatly assists the instrument's operational use. At a viewing distance of about 30 cm the test spot subtends an angle of 2° to the observer.

13. Calibration

In principle, the instrument produces an even reference field at Illuminant A and this is raised to Illuminant D_{65} by a $-196{\cdot}5$ mired (A to D) correction filter. The only unknown quantity in each instrument being the current the lamp should run at to produce Illuminant A. This could be found from previous lamp calibration and the stabilizer pre-set to a known current, but this raises problems of measurement accuracy when measuring the current, per lamp, under specified tested conditions, and later when re-establishing this current, quite probably, with different measuring equipment. Conversely, the A to D filter could be removed and the instrument pointed at a standard source of Illuminant A and the lamp current adjusted for a match, the A to D filter then being replaced.

The latter method makes the instrument nearly self-calibrating, particularly useful in view of the loss of blue light in the mirror box, as mentioned earlier. By increasing the lamp's colour temperature to offset this loss of blue light a pseudo-Illuminant A reference field is formed. The lamp life of 150 hours reported earlier was obtained under conditions necessary for this blue compensation.

There are then two recommended methods for calibrating the comparator, both of which rely on setting up a standard colour, viewing this with the comparator, and then adjusting the lamp current for a match in the comparator's matching field.

13.1 *Using a D_{65} Source*

If the instrument is pointed at a precise D_{65} source, then the instrument lamp can be adjusted for an optimum colour match with respect to both diffuser yellowing and any existing A to D filter errors. This method copies the accuracy of the test D_{65} source and is therefore recommended.

13.2 *Using an Illuminant A Source*

If the A to D filter is removed, the instrument can be pointed at an Illuminant A source and the lamp current adjusted for a match in the comparator's matching field. On re-inserting the A to D filter, the colour temperature of the reference field will be raised by the A to D filter together with any filter errors. This method is only recommended if other filters based on Illuminant A are to be inserted for other uses.

14. Stabilizing the Lamp Supply

The choice between stabilizing the lamp voltage or current is made weighing up the fact that a nominal change in current of say $1{\cdot}0\%$ has the same visual effect as a change in voltage of 2%, with the pros and cons of making a real stabilizer operate on voltage or current. At first sight it would appear that voltage stabilization would be preferred. However, this requires a measurement of voltage 'at the lamp' together with low resistance lamp socket connexions. If a stabilizer can be designed to a tighter tolerance for the current mode, the accurate monitoring of lamp voltage is avoided and advantage can be taken of the current mode to eliminate in-rush current and thereby promote longer lamp life. Current stabilization was therefore chosen for the calibrator.

14.1 *The Effect of Current and Voltage Variations on Colour Temperature*

Formulae offered for these calculations are approximations because lamps vary in performance according to their design and luminous efficiency. This has led to a small range of constants being quoted in various sources.[14–18] However, the order of difference between a change of voltage or current remains about 2 : 1.

We shall calculate the change in voltage and current that would produce a change of 2µd (2 mired) in colour temperature.

The relationship between luminous flux and colour temperature is given by a formula after Weaver and Hussong:[14, 18]

$$\log F = a_0 - \frac{11\,350}{T_c} \qquad \text{......(1)}$$

where F = luminous output, T_c = colour temperature, a_0 = constant.

The relationship between luminous flux and voltage or current is given by Walsh:[17]

$$\frac{\delta F}{F} = 3{\cdot}5\,\frac{\delta E}{E} \qquad \text{......(2)}$$

$$\frac{\delta F}{F} = 6{\cdot}5\,\frac{\delta I}{I} \qquad \ldots\ldots(3)$$

The 6 V, 10 W lamp used in the calibrator is rated at 200 lumens at 3200 K, giving 20 lumens/watt.

Substituting F in (1) to find the constant a_0 for this particular lamp:

$$\log 200 = a_0 - \frac{11\,350}{3200\text{ K}}$$

$$a_0 = 5{\cdot}851$$

A change of 2 µd at 3200 K is equivalent to 21 K. Adding this to 3200 and substituting for T_c in (1) will enable us to find the corresponding change in F:

$$\log F = 5{\cdot}851 - \frac{11\,350}{3221\text{ K}}$$

$$F = 214\text{ lumens}$$

The change in luminous flux for a colour temperature change of 2 µd is 14 lumens. This can now be substituted in (2) and (3) to find the corresponding change of voltage or current.

Substituting $\delta F = 14$ in (2) to find the change in voltage,

$$\frac{14}{200} = 3{\cdot}5\,\frac{\delta E}{6}$$

$$\delta E = 0{\cdot}12\text{ V}$$

The percentage change of the nominal value is

$$\frac{0{\cdot}12}{6} \times 100 = 2{\cdot}0\%$$

Substituting $\delta F = 14$ in (3) to find the change in current,

$$\frac{14}{200} = 6{\cdot}5\,\frac{\delta I}{1{\cdot}66}$$

$$\delta I = 0{\cdot}0179\text{A}$$

The percentage change of the nominal value is:

$$\frac{0{\cdot}0179}{1{\cdot}6} \times 100 = 1{\cdot}08\%$$

For the same change in lamp performance the voltage has to alter twice as much as the current, i.e. 2 : 1·08.

15. The Instrument in Use

Once the colour monitor calibrator has been switched on and an operator has handled it for half a minute or so, everything is really self-explanatory. Nevertheless a few words of instruction and explanation will help the operator to make a less ambiguous judgement.

Glare from white parts of a picture will cause poor seeing when using the instrument to look at dark areas. A hood big enough to prevent this would be considered too big operationally and would probably be left off! A better solution to glare is either to choose a test signal that does not produce any, or to cover part of the monitor with nylon sticker material to hide the offending area of glare. While making the match it is useful to wander the eye round the test spot edge, thereby reducing the build up of retinal fatigue.

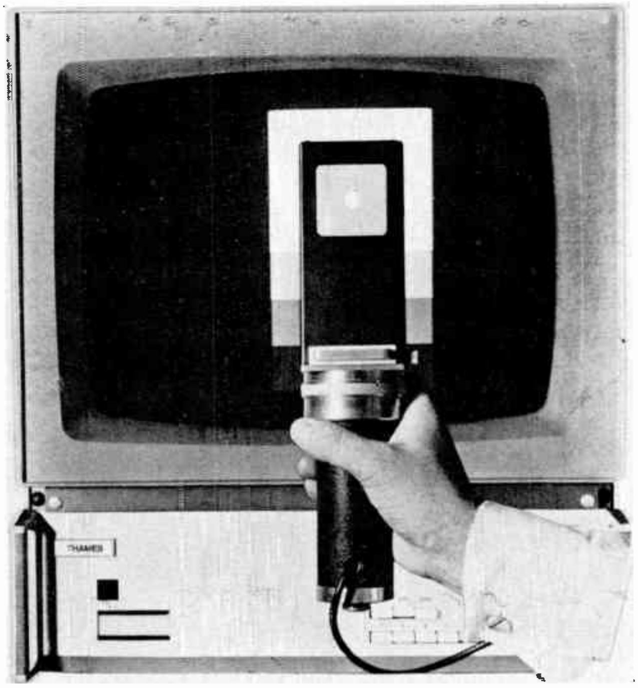


Fig. 2. An operator's view of the matching field when the instrument is held in front of a colour monitor. The reference field has been lowered in brightness, or mismatched, in order to show up the test spot and its surrounding reference field.

The iris brightness control allows the operator to exactly match the instrument's brightness to that of the colour monitor, thereby clearly isolating any chromaticity difference that may exist. The matching criteria then simply becomes one of chromaticity, and the operator

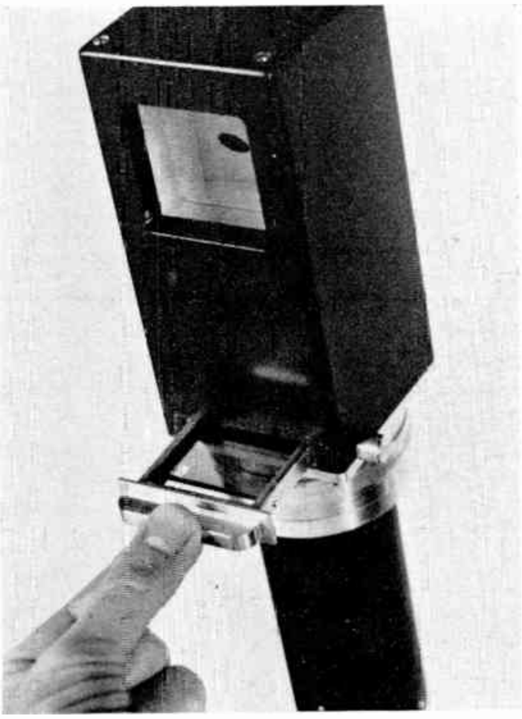


Fig. 3. A close view of the instrument illustrating the easy removal of the colour correction filter if another reference field colour is chosen.

ought to be able to arrive at a complete colour match—monitor controls allowing.

Because the test spot is televison raster light at 50 Hz, and because the instrument only sees a few lines of each field, the test spot has a very high off/on ratio. Sometimes the eye moves during the short on-period, upsetting the usual flicker integration process within the eye. When this happens part of the test spot momentarily blackens.

The instrument is not intended as a brightness standard. However, there is an adjustable stop on the iris in order that the peak brightness can be pre-set by the user: the instrument will match a screen brightness in excess of 30 ft-L (250cd/m^2).

16. Other Uses

The easy removal of the filter holder means the instrument can be quickly converted to check other colours. Various modes of colour monitor layout were discussed in the introduction and it is clear that in some cases a colour monitor will require a compromise white point setting other than D_{65}. For such instances the filter holder may be removed and a temporary gelatine filter (that a previous test has shown to be satisfactory) inserted.

It is possible to use the instrument for checking the colour reflected from film review screens. In this case another colour correction filter would be required for 5400 K instead of 6500 K.

Finally, specially computed filters up to 6 mm thick may be inserted to check the colour of shadow mask tube phosphors.

The production instrument† has the following outline specification:

Colour temperature:	6500 K $\pm 5\mu$d
Monitor brightness:	greater than 30 ft-L (250cd/m^2)
Brightness range:	continuous to black-out

17. Acknowledgments

The author thanks Bernard R. Greenhead, O.B.E., Director of Studios and Engineering, Thames Television Limited, for permission to publish this paper; Dr. Frank J. J. Clarke, N.P.L. for much help, advice, and the filter optimization calculations; and Alan Price, Grafikon, for successfully completing the onerous task of making the original idea work.

18. References

1. Knight, R. E., 'Preparations for colour tv', Brit. Kinematograph Sound Telev. Soc. paper, January 1969.
2. 'Colour Temperature for Colour Television Studio Monitors', S.M.P.T.E. Recommended Practice RP 37-1969.
3. Townsend, G. B., 'Transfer characteristics in the film—television process' (A proposed revision of the PAL System Signal Specification), B.K.S.T.S. Paper, September 1969.
4. Crawford, B. H., 'The Physiology and Psychology of Colour Rendering', National Physical Laboratory Symposium No. 8, 'Visual Problems of Colour', Vol. 2, pp. 383–385, H.M.S.O. 1958.
5. Wyszecki, G., and Stiles, W. S., 'Color Science', Sect. 5, 'Chromaticness and Brightness Matching'. (Wiley, New York, 1966.)
6. Newhall, S. M., Burnham, F. W., and Clark, J. R., 'Comparison of successive with simultaneous color matching', *J. Opt. Soc. Amer.*, **47**, No. 1, p. 43, 1957.
7. Read, B., and Hastings, W. W., 'An Ultraviolet Lighting Technique as used for Control Room Monitor Displays', Canadian General Electric Report, 31st May 1967.
8. Grum, F., 'Artificial light sources for simulating natural daylight and skylight', Proceedings of a Symposium on Colour Measurement in Industry, April 1967, pp. 109–122. (The Colour Group, Great Britain.)
9. McLaren, K. S., 'Illuminants for color matching', Proceedings of a Symposium on Colour Measurement in Industry, April 1967, pp. 81–104. (The Colour Group, Great Britain.)
10. Judd, D. B., 'Blue-glass filters to approximate the blackbody at 6500°K', *Farbe*, 10, p. 31, 1962.
11. Halliday, R. E. J.. 'The Optical Design of a Transparency Illuminator for Television Camera Alignment,' CBS Eng. Div. Report, 1965.
12. 'The Properties of Perspex Acrylic Materials' (I.C.I. Plastics Division).
13. 'Cementing of Perspex Acrylic Materials' (I.C.I. Plastics Division).
14. Wyszecki, G., and Stiles, W. S., *loc. cit.*, Sect. 1, 'Sources of Radiant Energy', p. 49.
15. 'I.E.S. Lighting Handbook', 'Incandescent filament lamp characteristics 8–11', 4th edition (Illuminating Engineering Society, New York, 1966).
16. 'Photometric Standards and the Unit of Light', N.P.L. Notes on Applied Science No. 24. (H.M.S.O., London, 1961.)
17. Walsh, J. W. T., 'Photometry', Appendix X and pp. 183–4. (Dover, New York, 1958.)
18. Weaver, K. S., and Hussong, H. E., 'A note on the colour temperature-candle power characteristic of tungsten lamps', *J. Opt. Soc. Amer.*, **29**, p. 16, 1939.
19. Smith, W. J., 'Modern Optical Engineering'. (McGraw-Hill, New York, 1966.)
20. Judd, D. B., 'Sensibility to colour temperature change as a function of temperature', *J. Opt. Soc. Amer.*, **23**, No. 7, pp. 7–14, January 1933.

Manuscript first received by the Institution on 20th February 1970 and in final form on 7th April 1971. (Paper No. 1397/Comm 49.)

† Manufactured by Grafikon (Engineers) Ltd., 75 South Western Road, Twickenham, Middlesex.