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 50 No. 4 April 1980

The Radio and Electronic Engineer

The Journal of the Institution of Electronic and Radio Engineers

Engineering under the Microscope

TWO years ago the 1977 Survey of Professional Engineers reported* that incomes of Chartered Engineers had fallen further in comparison with industry since the previous survey: it will come as no surprise to IERE members that the 1979 Survey† shows that the median incomes of engineers have fallen in real terms since 1976/7 whereas the real average earnings index for all industries showed a significant increase. Indeed, in relation to the median income for all Chartered Engineers of £7200, p.a. electronic and electrical engineers were generally even below that figure by a hundred or so pounds especially those in productive, industrial fields of work. Moreover they were between £1000 and £1500 a year worse off than their colleagues in 'service' industries such as Telecommunications and Broadcasting or in the Universities, and even those in Computer Technology were on the median annual income of £7200. In this year of the Finniston Inquiry into Manufacturing Industry, and in the 'Golden Age' of electronics, the rewards for those responsible for creation of wealth seem even more niggardly than those of the profession as a whole.

The Survey presents much fascinating information relating to the background, deployment and employment of engineers, and some of the points are discussed in a commentary by the Chairman of the Steering Committee concerned with its preparation, Mr Bryan Hildrew, C.B.E., F. Eng. He suggests that if it is true that this country is dependent upon exporting technology to survive, then it needs high quality engineers. The survey must surely alert all to the fact that however excellent the quality of engineers produced by our Universities and Training Establishments, they will not enter or remain in a profession which is subjected to such an erosion of its financial rewards.

Graphs of Distribution of Income indicate that it is only after reaching the age of 45 that the most competent and effective of the engineering profession, the highest decile, achieve markedly higher incomes. It is probable that this upturn of the curve owes a great deal to income from consultancy services in an international context.

Historically the comparison of Chartered Engineers non-graduates', graduates' and post-graduates' income against age has been remarkably consistent and the rewards accorded to those with higher qualifications have been decreasing. Mr Hildrew notes that post-graduates, particularly in the upper quartile, have now established a considerable differential over the age of 55. The sample is a considerable one and again it is possible that consultancy fees are the prime cause.

Two-thirds of all Chartered and two-thirds of all Technician Engineers‡ now work a contractual working week of 38 hours or less. However, over 40% in each category put in over 42 hours per week. Thirty per cent of Chartered Engineers and 43% Technician Engineers receive payment for overtime, an increase over the 1977 recorded figures of 20% and 38% respectively. Rather surprisingly, a small number of engineers work up to 25 hours per week overtime and the higher the salary earned the greater the hours of overtime, paid or unpaid, worked.

The analysis of qualifications shows that two-thirds of the Chartered Engineer population now qualify educationally by degree equivalent. This large graduate identity which is growing steadily, together with the post-graduate population, will, in Mr Hildrew's opinion, provide a sound base when negotiating equality of professional status within the European Economic Community.

There has been a further slight increase in the proportion of Chartered Engineers in the Public Sector possibly due to Aerospace and Shipbuilding being taken into public ownership. There is also an increase in the numbers of Chartered Engineers employed in industrial or commercial companies and a decrease in the self-employed and those serving in the Armed Forces.

The analysis by occupation shows an increase in those Chartered Engineers employed in Research & Development and in Design. In the case of Design it is to be hoped that the continuous decline in this occupation since 1966 has at least been halted. A similar upward trend is recorded for Technician Engineers. These two occupations have low average ages and associated low salaries. It is necessary to ensure by financial incentive that the more competent Designers and R & D personnel remain in these essential occupations. Good R & D and Design are vital investments in the nation's future.

The analysis of the subjective question relating to responsibility indicates clearly that responsibility is in fact

* 'The 1977 Survey of Engineers', *The Radio and Electronic Engineer*, **49**, page 96. January/February 1979

† 'The 1979 Survey of Professional Engineers". Published by CEI, Little Smith Street, London SW1P 3DL. Price £5·20 including posting and packing in the UK.

‡ The Technician Engineer Survey, which was based numerically on about a tenth of the 25 000 completed questionnaires from Chartered Engineers is heavily weighted towards the electrical and electronic field as there is no widely accepted mechanical institution catering for the Technician Engineer.

rewarded and that it is related to age. Respondents were asked to describe their own jobs in terms of six factors:—

technical complexity, commercial responsibility, physical hazard, value of equipment used, absence from home, experience.

The six factors all tend to contribute to earning a higher salary at almost every age band. However, the power of the various factors is very different.

The pattern of remuneration by age for those whose work is scored 'high' and 'low' on each of the six factors shows that in every case, incomes tend to rise with age except for the final age group. It is very striking that technically complex jobs are often paid less than technically simple ones.

Commercial responsibility earns a larger and larger income differential up to the early 50's, after which the differential stabilizes— at almost £2,000 per annum.

Physical hazard is generally compensated by a differential of £300–£400 per annum, while responsibility for expensive equipment earns a rather larger differential of over £400 per annum. Absence from home earns the largest differentials, though interestingly, there is a sharp change at age 30. Below this age, only small differentials less than £200 per annum are earned—perhaps because single people need less compensation for working away from home than do married. The differentials for those aged more than 30 exceed £1,000 p.a. and rise to more than £5,000 for those aged 60 or more, Experience does not seem to earn a large differential for a given age band, though overall those with 'high' experience are paid £1,300 per annum more than those with 'low'. This may be because respondents found it difficult to distinguish between the concepts of age and experience.

The mobility of engineers has been linked to the location of employment. It will be noted that the degree of mobility with current employer or by change of employer over the chosen periods of 2 years and 6 years is remarkably even over the country and has a degree of proportionality in relation to the two time periods. The two major exceptions are Northern Ireland, where mobility is restricted by virtue of the ongoing emergency, and the Republic of Ireland where the limited sample size distorts the figures.

The tabulation of mobility against age not unexpectedly shows that after the age of 44 the number of moves, particularly to new employers, decreases rapidly. Mobility has also been related to field of work and, overall, the Chartered Engineer is shown to be more mobile than the Technician Engineer. Of particular note is the high degree of mobility demanded of both Chartered and Technician Engineers in the Armed Forces.

It is noteworthy that the percentages of Chartered Engineers and Technician Engineers who are members of a Trade Union have remained virtually constant since 1977. Change in membership distribution between unions is very limited.

The method of determination of salary has been related to the Private and Public employer by Institution. It is apparent that where collective bargaining is done from strength as it is in the public sector, it is effective. When done from a position of weakness its effectiveness is limited. Personal representation or Periodic Adjustment of salary would appear to achieve similar or better results in both the Public and Private Sectors.

Once again, pensions and provision of a company car were compared. The number of Chartered and Technician Engineers who had no pension scheme in 1977 has been more than halved. There has been an increase in the numbers enjoying index-linked pension schemes, particularly including Technician Engineers, and about 20% of all engineers in the Private Sector state that they have index-linked pensions.

Surprisingly, provision of a company car is still restricted to only 20% of the profession, a figure identical to that obtained in the 1977 Survey. The distribution is predominantly to Senior Chartered Engineers. whereas Senior and Junior Technician Engineers enjoy a wider distribution.

There has been an increase in training in Technology and in Management which counters the decline commented upon in 1977. The proportion of the respondents in receipt of further training is higher for Technician Engineers and covers a wider age spectrum.

The number of Chartered Engineers unemployed continues to fall and the duration of unemployment is concentrated in short-term (up to 12 weeks) which must contain a proportion of engineers changing employment and in long-term (over 32 weeks) which contains those few engineers having great difficulty in obtaining employment.

Mr Hildrew considers that, in total, this Survey reinforces the particular point arising from the 1977 Report. The United Kingdom has a very stable and coherent professional engineering workforce. The educational qualifications for both Chartered and Technician Engineers are clearly identifiable and represent the foundation on which the profession is firmly established.

His conclusions with which few will quarrel are that the nation's technical standing in the world market must be stimulated, restored and recognized in order to maintain and if possible improve the quality and fabric of life in this country. Therefore it is fundamental that the decline in rewards for the professional engineer's contribution to the economy be reversed.

F.W.S.

Applicants for Election and Transfer

THE MEMBERSHIP COMMITTEE at a meeting on 11th February 1980 recommended to the Council the election and transfer of the following candidates. In accordance with Bye-law 23, the Council has directed that the names of the following candidates shall be published under the grade of membership to which election or transfer is proposed by the Council. Any communication from Corporate Members concerning the proposed elections must be addressed by letter to the Secretary within twenty-eight days after publication of these details.

February Meeting *(Membership Approval List No. 269)*

GREAT BRITAIN AND IRELAND

CORPORATE MEMBERS

Transfer from Graduate to Member

COFFEY, Thomas B. *Enfield, Middlesex.*
ROBERTS, John Richard. *Bromley, Kent.*
ROBERTSON, Graham. *Wyton, Hunts.*
SOOD, Subash Chander. *West Bridgford, Notts.*
SQUIRES, Peter John. *Erith, Kent.*
THOMAS, Stanley Richard. *Weymouth, Dorset.*
WONG, Andrew Chi-Chung. *Harlow, Essex.*

NON-CORPORATE MEMBERS

Transfer from Student to Graduate

PARTAP, Robert Kamal. *Plymouth, Devon.*

Direct Election to Graduate

BHANDARI, Narinder Singh. *Wembley, Middlesex.*
HOPPER, Simon John. *Guildford, Surrey.*
HUGILL, Timothy John S. *Porthcurno, Cornwall.*
INIGHT, Keith Paul. *Salihull, West Midlands.*
MARTIN, Anore Keith. *Porthcurno, Cornwall.*
ORMOND, David. *Porthcurno, Cornwall.*
WATKINS, Kevin. *Porthcurno, Cornwall.*

OVERSEAS

CORPORATE MEMBERS

Transfer from Graduate to Member

CHAN, Yan Lam. *Pokfulam, Hong Kong.*
CHEUNG, Leung Fat. *Wanchai, Hong Kong.*
SEE, Ah Bah. *Singapore.*
WELTON, Peter John. *Willoughby, New South Wales.*

NON-CORPORATE MEMBERS

Direct Election to Graduate

CHU, Man To. *Hong Kong.*
IP, Hay Shing. *Kennedy Town, Hong Kong.*

Direct Election to Associate Member

HARDY, David. *Tripoli, Libya.*
MUHAMMAD, Ajmal Jauhar. *Naushahra, Pakistan.*
WONG, Pak Yeong. *Shah Alam, Selangor, West Malaysia.*

Direct Election to Student

CHAN, Kam Hung. *Kowloon, Hong Kong.*
CHIANG, King Wah. *Kowloon, Hong Kong.*
HO, Kam Wing. *Kowloon, Hong Kong.*
HUI, Wing Hong. *Kowloon, Hong Kong.*
KWOK, Sui Ping. *New Territories, Hong Kong.*
SMITH, Peter. *BFPO 20.*

Direct Election to Associate Member

BLACKWELL, Brian Richard. *Colchester, Essex.*
COX, Peter Kevin R. *Alnwick, Northumberland.*
CRIPPS, John Edwin. *Oswestry, Salop.*
HEARN, Peter Alexander. *London.*
HIND, Gordon Wilson. *Tynemouth, Tyne & Wear.*
HURLEY, Patrick de Sousa. *Maldon, Essex.*

Direct Election to Student

FARMER, Mark. *Bath.*
FOYLE, Stephen Peter. *Reading.*
GILL, Deshinder Singh. *Oxford.*
JAMES, Stephen William. *Swansea.*
KING, Paul David. *Bath.*
OLLEY, Christopher Andrew. *Bordon, Hants.*
PANNETT, Stephen John. *Swansea.*
PASELLA, Dean. *Swansea.*
RAYIT, Tejpal Singh. *Canterbury, Kent.*
ROWE, Timothy George. *London.*
SANDHU, Surjit Singh. *Bath.*
ZDROJKOWSKA, Kristyna. *Newcastle-upon-Tyne.*

Conference on 'Digital Processing of Signals in Communications'

The continued rapid advances in Digital Circuit Technology over the past few years have led to more and more complex devices being available at relatively low cost and of extremely small size. This advance has influenced a number of other fields causing them to consider systems which would have been uneconomical with discrete components. The communications field is no exception and many new systems are being introduced as a direct result of the availability of medium and large scale integration devices and in particular the microprocessor. These advances in the Communications and Television field form the theme of the present conference, to be held at the University of Technology, Loughborough, from 6th to 10th April 1981. It is the second IERE conference on this theme to be held, the first being in Loughborough in 1977. Co-sponsors of the conference are the Institution of Electrical Engineers and the Institute of Electrical and Electronics Engineers.

The Organizing Committee, which is under the Chairmanship of Professor J. W. R. Griffiths, and includes representatives from Industry, Government Research Establishments and Universities, now invites papers for this conference. In the first instance synopses of proposed contributions are requested. Preference will be given to those describing original work, new systems, or new applications.

The Committee would also like to see papers on practical aspects of communications systems. For guidance a non-exclusive range of subjects is offered:

antenna array signal processing
coding techniques for television transmission
computer processing of images
control of distributed processing
digital filters in communication
digital processing of speech and music signals
digital studio techniques
echo cancellation
economics of signal processing
error detection and correction
high speed modems
noise reduction techniques
programmable processing
signal design
simulation of communications systems
spread spectrum techniques
telemetry
underwater acoustic communications
use of computer array processors

A synopsis, of not more than 300 words, will enable the Committee to assess the scope of the proposed paper, and should be sent to the Conference Secretariat as soon as possible but not later than 1st June 1980. Papers will be required in their final form by 1st November 1980.

Further information and registration forms for the conference will be available in due course from the Conference Secretariat, Institution of Electronic and Radio Engineers, 99 Gower Street, London WC1E 6AZ (Tel: 01-388 3071).

Members' Appointments

CORPORATE MEMBERS

F. P. A. Allin (Member 1969, Graduate 1964) has been appointed Product Engineer with Fuchs Electrical Industries, Transvaal, South Africa.

M. O. Anyiam, M.Sc. (Member 1974, Graduate 1963) has been appointed Rector of the Federal Polytechnic, Bauchi, Nigeria. Mr Anyiam, who obtained the M.Sc. in microwave physics from the University of Surrey in 1966, was formerly with the Nigerian Broadcasting Corporation.

I. E. Dunn, B.A. (Member 1969) has taken up the position of Assistant Head of Engineering Training with the Thompson Foundation Television College in Newton Mearns, near Glasgow. Mr Dunn who joined the BBC in 1960, was latterly Principal of the Technical Training Centre of the National Broadcasting Commission in Papua New Guinea.

Sqdn Ldr M. D. L. Mason, B.Sc. (Econ), RAF (Ret.) (Member 1973) is now Managing Director of Technical Translation International, based in London.

R. F. Mason (Member 1966), formerly terotechnology consultant with Plessey Assessment Services, has taken up an appointment as Professional Officer to the British Orienteering Federation.

E. N. Maynerd (Member 1967) who was Head of the Aeronautical Engineering Division of AQD in the Procurement Executive of the Ministry of Defence, has now retired from Government Service.

U. D. Munasinghe (Member 1978, Graduate 1973) has been promoted to Superintendent Telecommunications Engineer with the Posts and Telecommunications Department, Government of Sri Lanka. He has been with the Department since 1964 when he completed full-time studies at the Institute of Higher Technology, Katubedde, Sri Lanka.

D. O. U. Nwachukwu (Member 1974) is now Head of Telecommunications for Shell Petroleum Development in Lagos. He joined the Company as a junior technician engineer in 1963 after obtaining the Diploma in Communications and Electronics of Southampton College of Technology.

S. E. Osime (Member 1974, Graduate 1971) has had his appointment as Principal Engineering Manager with the Nigerian Posts and Telecommunications Department in Benin City confirmed.

R. J. Preston (Member 1971, Graduate 1966) has been appointed a Research Director in Racal Research. He has been a senior engineering executive with Racal Mobilcal, having joined the Group in 1967.

K. R. Thrower (Member 1967, Graduate 1965) has been appointed Managing Director of a new company within the Racal Group, Racal Research. Since 1974, Mr Thrower has been Director of Advanced Development with the Group which he joined in 1957. A member of the Institution's Communications Group Committee since 1974, and of various conference organizing committees, he was elected to Council at the last Annual General Meeting.

NON-CORPORATE MEMBERS

P. J. Mullis (Graduate 1967), having retired from REME in 1977 with the rank of Captain, is now Design Assurance Manager with Plessey Telecommunications Systems at Poole.

Pang Ka Foh (Associate Member 1977) who has been an Aircraft Radio Technician with the Royal Malaysian Air Force since 1962, has now joined Siemens Components, Batu Berendam, Free Trade Zone, Malacca, as a Section Supervisor.

A. J. Towning (Graduate 1971) is now with London Air Traffic Control as a software engineer.

Obituary

The Institution has learned with regret the deaths of the following members

Dennis Baxter (Associate Member 1979, Associate 1951) of Preston, died recently aged 68. Prior to his retirement due to ill health in 1975, Mr Baxter was a lecturer on the staff of Harris College of Further Education (now Preston Polytechnic).

Allan Geoffrey Davies (Member 1973, Graduate 1966) of Havant died on 6th October 1979 aged 44 years; he formerly held appointments with Plessey Product Assessment Laboratories and with Bradma.

Keith Christian Davis (Member 1973) died on 9th December 1979 at the age of 52 years. He had held the appointment of Deputy Controller as Head of the Telegraph Mechanisation Engineering in the Post Office International and Maritime Telecommunication Region since 1971.

Geoffrey Huntley Dawe (Member 1964, Graduate 1959) died last autumn, aged 63. He was formerly a Test Engineer with the Electricity Department of the City of Salisbury, Rhodesia.

Douglas Cecil Dix (Graduate 1967) of Wembley, an Executive Engineer with the Post Office, died last year, aged 41 years.

Christopher Brian Hames (Member 1970) of St. Albans died in January 1980, aged 46. He was formerly a Senior Design Engineer with Hawker Siddeley Dynamics.

Ernest John Hudgell (Member 1973, Graduate 1963, Associate 1948, Student 1943) of Farnham died in November 1979, aged 61. Since 1963 he had been on the Staff of Farnborough Technical College, latterly as a Lecturer II teaching final year courses for City and Guilds examinations on computers.

James Mason (Member 1974, Graduate 1970) of Northallerton, died in September aged 49. He had held appointments with the Newcastle Regional Hospital Board as a Main Grade Engineer and as an Electrical Engineer with the North Yorkshire County Council.

Alfred George Paracelsus Mower, B.Sc. (Fellow 1964, Member 1941) died on 18th December 1979 aged 67 years leaving a widow. After graduating with a London University degree in physics, Mr Mower joined Callenders Cable and Construction Company in 1933 as a junior engineer in the Measurements Laboratory. During the war years he was with the Ministry of Aircraft Production, as an inspector of radio equipment and after a short period with Taylor Electrical Instruments in 1946/47, he joined Philips Mitcham Works with whom he remained until his retirement due to ill health in 1971. He was concerned throughout with radio and television design and measurements and as well as serving on industry committees for the company and on an IEC Committee, he represented the Institution on BSI Technical Committees. For some eight years he was a member of the Institution's Examinations Committee.

Alfred Lewis Trimmer (Member 1961) of Stevenage died on 12th April 1979, aged 59 years.

Conference Report

IMACS Congress 1979: Simulation of Systems

Organized by Hybrid Computers Research Centre of the National Research Council and University of Naples for the International Association for Mathematics and Computers in Simulation and held in Sorrento, Italy, from 24th to 28th September 1979.

This ninth congress was well attended and included a strong UK contingent of simulation specialists. The location and weather were particularly pleasant, but hardly conducive to full attendance at all the sessions of presented papers. This was allowed for by the organizers and technical visits were arranged to the Olivetti-Osai factory in Casserta, for demonstration of digitally-controlled machine tools and robots, and to the Hybrid Computers Research Centre at the University of Naples. As relaxation all participants were invited to a half-day boat trip, followed by refreshment on the Isle of Capri. In addition, most participants took advantage of the unique tourist attractions of the area and made visits to the ancient city of Pompeii and the quiescent volcanos of Solfarata and Vesuvius.

As indicated by the title the dominant theme was simulation and the aims were to present recent trends in the fields of modelling and simulation and to make a survey of the actual applications of simulation of systems. To judge by the presented papers these aims were certainly achieved. One would believe that every type of engineering system had already been simulated and that complex biological and social systems were now under active consideration.

The congress topics were divided into three main areas:

Methodology in Modelling and Tools.
Simulation of Well-Defined Systems.
Simulation of Ill-Defined Systems.

The pattern of presentation for each area was to have an invited lecture by an international expert, together with a large number of relevant contributed papers. The latter papers were presented in three parallel sessions, corresponding to the above divisions, throughout the week.

The first area covered simulation methods, socio-economic systems, population dynamics, data processing systems, and simulation software. B. P. Ziegler gave the first invited lecture 'Modelling and simulation methodology, state of the art and promising directions'.

Engineering topics were dominant in the second area which covered electrical, mechanical, civil, control, thermal and nuclear. As an example of a control engineering topic, the writer presented a paper, 'Dynamic simulation of water distribution systems'. This paper outlines the simulation requirements for on-line computer control of water systems.

The radio and electronic engineering aspects were not neglected and occupied a full morning session with six presented papers on simulation topics, ranging from communication systems to integrated circuits. A typical paper is by J. W. Modestino, *et al.*, 'Digital simulation of communication systems' which describes a hardware-software system for the digital simulation of arbitrary point-to-point communication systems. Of particular interest in this second area were the digital computer exhibits loaned by the congress organizers and available for interactive use by the participants.

Biological and socio-economic systems came under the third area of ill-defined systems. The invited lecture to be given by R. Tomovic, 'Simulation of ill-defined systems', was anticipated to provide more definition in this difficult area. Unfortunately, the author was unable to attend and L. Dekker was called upon at short notice to present a special paper 'Views on parallel simulation'.

The closing event of the congress was a special session of invited papers discussing the state of the art and future trends with emphasis on parallel processing in relation to parallel simulation. In order to meet the requirements for simulation of large-scale systems which may be non-linear and ill-defined, it was expected that advances would be made in the following areas:

(i) Simulation Hardware—involving development of parallel processing capability to allow high-speed simulation using parallel computing technique.
(ii) Simulation Software—involving development of more general, user-oriented, simulation language able to cope with an increasing variety of non-standard situations in a systematic manner.

The diverse range of simulation applications and the necessary parallel sessions of contributed papers inevitably gave rise to small audiences. Fortunately this did not apply to the sessions of invited lectures, dealing with more general simulation techniques, which usually attracted capacity audiences. It will be desirable to question the usefulness of future simulation conferences encompassing such a wide range of topics. As an overall view it is apparent that systems simulation is now a major subject area in its own right, rather than just a useful descriptive convenience for engineers and scientists. The title 'Simulation Engineer' could well be on the way.

B. COULBECK

The complete Conference proceedings 'Simulation of Systems' are to be published by the North-Holland Publishing Company, Amsterdam.

The Science Research Council Reports...

In its fourteenth Annual Report* the Science Research Council reviews its financial situation and provides a summary of its plans. Particular attention is drawn to the serious effect the lack of career opportunities for young scientists in universities and polytechnics is beginning to have on the development of new science and the difficulties universities face in finding funds to re-equip laboratories.

The Report reviews the substantial progress in the special programmes in the engineering fields to which the Council attaches high priority—Marine Technology, Polymer Engineering and the Teaching Company—and the initiatives it has taken in support of microcomputer technology, software and robotics. It emphasizes the importance it attaches to the Council-wide scheme of grants for collaborative research programmes with industry. For example, over £1M of proposals have already been received from a wide range of industries and first grants made in the Co-operative Research Scheme launched during the year and new education and training schemes (the Collaborative Training Award Scheme and the Ex-part-time Scheme) were introduced.

The Council has been concerned that the high rate of technological development means that the scientific knowledge of many graduates in industry is out of date. A pilot programme of post-experience training for those experienced in industry but needing to be updated in technological advances is likely to involve the Open University in co-operation with other academic institutions. As a forerunner to a scheme for new graduates, the Council is going to support a limited number of bursaries at some polytechnics.

Progress on the new central facilities, in particular the Nuclear Structure Facility, the Synchrotron Radiation Source, the Spallation Neutron Source and the High Power Laser equipment, is charted, the Council recognizing how essential access to advanced research facilities is if UK scientists are to stay in the forefront of their fields.

The accompanying bar chart shows how the actual domestic and international expenditure of the Science Research Council has been distributed between the various Boards and the other facilities in the years 1973/74 and 1978/79 together with the similar proposed expenditure for 1983/84. Under the heading of international bodies and projects are included the European Organization for Nuclear Research, European Space Agency, Institut Laue-Langevin, Anglo-Australian Telescope, South African Astronomical Observatory, European Incoherent Scatter Scientific Association and North Atlantic Treaty Organization.

** Report of the Science Research Council for 1978–79. Published for SRC by HMSO. Price £5.25 net.*

SRC's special schemes

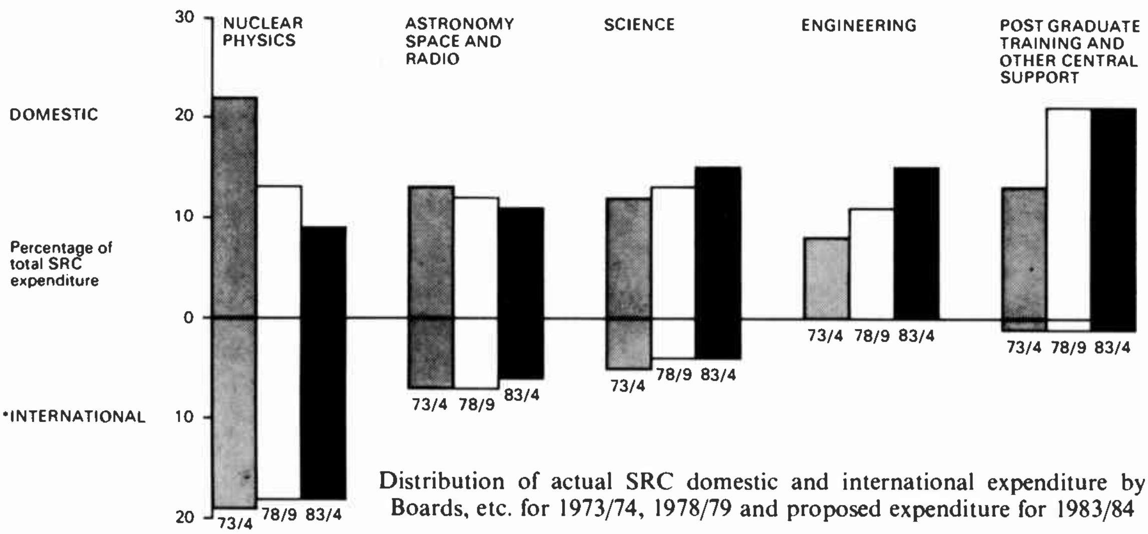
A series of leaflets has been produced by the Science Research Council to aid the promotion of some of the special schemes run in co-operation with industry and other collaborating bodies. These schemes have been specially devised to broaden the scope and effectiveness of the postgraduate training available in areas of national importance.

The schemes provide awards to enable postgraduate students to gain experience of working with or in industry (Co-operative Awards in Science and Engineering, Collaborative Training Awards, Total Technology) or they can give young graduates who have been working in industry the chance to undertake full-time postgraduate training (Instant Awards, Full Time Training for Former Part Time Students).

Joint SRC/Social Science Research Council Studentships are offered for taught courses and research programmes that explore the interplay between technical and social issues.

The Teaching Company scheme, run jointly with the Department of Industry, operates differently: its aim is to develop active partnerships between universities/polytechnics and manufacturing companies through programmes in which the academic partner helps the company to achieve some major change in manufacturing. Teaching Company associates, appointed by the academic partner in consultation with the manufacturing company, are normally based full time at the company's works and undertake tasks within the programme. Their salaries and academic support costs are paid by SRC/DoI grant.

More leaflets are planned in the series, seven of which are available in this first batch. They are available free of charge from the Press Office, Science Research Council, PO Box 18, Swindon SN2 1ET (Tel: 0793–26222, ext 2256/7).



Distribution of actual SRC domestic and international expenditure by Boards, etc. for 1973/74, 1978/79 and proposed expenditure for 1983/84

New Books Received

The following books which have been received recently have been placed in the Institution's Library and may be borrowed by members resident in the British Isles.

Hybrid Microelectronics
T. D. TOWERS. Pentech Press, London; Crane, Russak & Company, New York, 1977. 22 × 14 cm. 246 pages. £7.50.
CONTENTS: Hybrid microcircuits. Hybrid microcircuit fabrication. Special user features of hybrids. Hybrid microcircuit manufacturers. Hybrid microcircuit packages. Passive hybrid microcircuits. Motherboard hybrid microcircuits. High power hybrid microcircuits. Precision hybrid microcircuits. R.f. hybrid microcircuits. Optoelectronic hybrid microcircuits. Home entertainment hybrid microcircuits. Hybrids in domestic appliances. High-reliability hybrid microcircuits. Instrumentation hybrid microcircuits.

The 8080A Bugbook. Microcomputer Interfacing and Programming
PETER R. RONY, DAVID G. LARSEN and JONATHAN A. TITUS. Howard W. Sams, Indianapolis, 1977. 21 × 13 cm. 408 pages. £6.95.
CONTENTS: What is a microcomputer. A small 8080 microcomputer. An introduction to microcomputer programming. Generating a device select pulse, clock cycles and timing loops. Generating status information. Microcomputer input/output. Subroutines, interrupts, external flags and stacks.

Modern Dictionary of Electronics
RUDOLF F. GRAF. Prentice-Hall International, Hemel Hempstead, 1977. 22 × 14 cm. 832 pages. £13.25.
While this volume is as up-to-date as possible at the time of writing, the field of electronics is expanding so rapidly that new terms are constantly being developed and old terms are taking on broader or more specialized meanings. It is the intention of the publishers periodically to issue revised editions of this dictionary; thus they welcome suggestions for new terms, and definitions.

Solid State Devices 1977
Edited by E. A. ASH (*University College London*). Institute of Physics, London, 1977. 24 × 16 cm. 175 pages. £16.50.
CONTENTS: Future directions of silicon device technology. Silicon-on-sapphire device: realization, properties and applications. Problems related to avalanche multiplication in p-n junctions and Schottky barriers. Fine-particle magnetic materials. Electrically controlled colour filters using liquid crystals. High-reliability transistors for submarine systems. Analogue signal-processing devices. Microwave field effect transistors. A review of high-resolution microfabrication techniques.

Tape Recorder Servicing Manual
JOHN GARDNER. Newnes-Butterworths, Sevenoaks, 1977. 25 × 19 cm. 205 pages. £7.80.
CONTENTS: Test equipment. Performance tests. The Dolby system. Notes on dimensions.

Newnes Colour TV Servicing Manual—Vol. 3
GORDON J. KING. Newnes-Butterworths, Sevenoaks, 1977. 19 × 25 cm. 233 pages. £7.80 (hardcover).
CONTENTS: Rank-Bush-Murphy Z179 chassis. Hitachi CSP-680 receiver. ITT CVC8 chassis. Bang & Olufsen Beovision 4000 and 5000 receivers. Decca solid state 40 series receivers. Thorn 9000 series chassis. Philips G9 chassis.

Microprocessors and Small Digital Computer Systems for Engineers and Scientists
GRANINO A. KORN (*Professor of Electrical Engineering, University of Arizona, Tucson*). McGraw-Hill, Maidenhead, 1977. 23 × 16 cm. 380 pages. £18.40.
CONTENTS: The ingredients of mini-micro-computation. Processors and instruction sets. Programming with assemblers and macro-assemblers. Interfacing microprocessors and minicomputers with real-world devices. Computer operation, software, and program preparation. Microcomputer systems and design decisions. Micro-minicomputer families.

Handbook of Linear Integrated Electronics for Research
T. D. S. HAMILTON (*Senior Lecturer at the University of Manchester*). McGraw-Hill, Maidenhead, 1977. 25 × 19 cm. 450 pages. £14.50.
CONTENTS: Review of basic circuit theory. Operational amplifiers. Stability of feedback systems. Amplifiers. Oscillators. Circuit functions. Power supplies. Circuit devices. Optoelectronics. Signal detection. Applications.

Electronics. TEC Level II
D. C. GREEN (*Senior Lecturer in Telecommunication Engineering at Willesden College of Technology*). Pitman Publishing, London, 1978. 24 × 19 cm. 150 pages. £3.50.
CONTENTS: Simplified semiconductor theory. Semiconductor diodes. Transistors. Thermionic valves. The cathode ray tube. Small signal amplifiers. Waveform generators. Digital elements and circuits. Power supplies.

Heinrich Hertz—Memoirs, Letters, Diaries
JOHANNA HERTZ (Second edition edited by Mathilde Hertz and Charles Susskind). San Francisco Press, 1977. 23 × 15·5 cm. 344 pages. $19.50.
This biographical compilation is as fascinating as a picture of Germany in the latter half of the nineteenth century as for the account it gives of the life work of the first experimental radio scientist. The translation and editing, by Hertz' younger daughter and Professor Charles Susskind, of the University of California, who is a member of the IERE, can hardly be faulted for its smoothness and felicity; its fidelity can be checked against the original German texts with which it is interleaved.

Soft-soldering Handbook
C. J. THWAITES (*Tin Research Institute*). International Tin Research Institute, Greenford, Middlesex (ITRI Publication 533) 24 × 15 cm. 102 pages. £3.00.
CONTENTS: Principles of soft-soldering. Design and materials for soldering. Soldering methods. Testing and quality control. Safety-Health and environmental aspects.

Motion Picture Camera and Lighting Equipment
DAVID W. SAMUELSON. Focal Press, London, 1977. 14 × 21 cm. 220 pages. £2.95 (paperback).
CONTENTS include: Camera movement. The shutter. Adjustable shutters. Viewfinder systems. Aperature plate. Lens mounting. Lens/camera compatibility. Lens assessment.

Introduction to Communication, Command and Control Systems
DAVID J. MORRIS (*Senior Lecturer in the Electrical Engineering Department at the Ben-Gurion University*). Pergamon Press, Oxford, 1977. 25·5 × 18 cm. 308 pages. £15.00 (hardcover).
CONTENTS: System design concepts. Sensor base data collection. Data transmission. Multiplexers and concentrators. Switching centres. Communication network hierarchy, architecture and control. Loop transmission. Computers in command and control systems. Distributed computer resources. Terminals and displays. Error control. Secrecy, security and privacy. Reliability and maintainability.

Electric Circuit Theory
J. M. IVISON (*Department of Electronic and Electrical Engineering, Loughborough University of Technology*). Van Nostrand-Reinhold, New York, 1977. 30 × 21·5 cm. 118 pages. £10.00 (cloth), £4.50 (paperback).
CONTENTS: Circuit elements and sources. Some circuit theorems and their application to d.c. circuits. Methods of circuit analysis. The transient response of simple circuits. The application of the Laplace transform to electric circuits. Circuits with sinusoidal excitation. Voltage and current relationships in a.c. circuits. The steady-state frequency response of circuits. Linear two-port networks. Locus diagrams. Three-phase circuits.

Programming for Minicomputers
J. C. CLULEY (*Department of Electrical and Electronic Engineering, University of Birmingham*). Edward Arnold, London, 1978. 23·5 × 16 cm. 262 pages. £12.50.
CONTENTS: Computing environment. Arithmetic and logical operations. Machine mode instructions. The use of assembly languages. Multiple register and multiple address. Basic programming techniques. Imput-output instructions. Microprocessors and their order codes. System programs. On-line programming.

Video Tape Recorders
HARRY KYBETT (*Sony Corporation of America*). Howard W. Sams, Indianapolis, 1978. 21 × 13 cm. 394 pages. £7.25 (paperback).
CONTENTS: Brief history of video recording. Review of audio recording. Principles of video recording. The broadcast quad-head recorder. The helical vtr. The mechanics of helical vtrs. The record and playback system. Servos. Quad servos. Control pulses and other functions. Editing color recording and playback. Time-base errors and their correction. Video cassette machines. Broadcast helical vtrs. The portable vtr. Digital television.

Modern Switching Theory and Digital Design
S. C. LEE (*Professor of Electronic Engineering and Computing Sciences, University of Oklahoma*). Prentice-Hall, Englewood Cliffs, N.J. 1978. 18 × 24 cm. 490 pages. £16.75.
CONTENTS: Boolean algebra and Boolean function. Vector switching algebra and vector switching function. Boolean differential calculus. Special switching functions. Multivalued logic. Fault detection in combinational circuits. Sequential machines. Regular expressions. Sequential machine realization. Sequential machine fault detection. Digital design using integrated circuits. Digital design using microprocessors via simulation.

Audio Technology Systems
DEREK CAMERON. Prentice-Hall International, Reston, Va. 1978. 23 × 13·5 cm. 245 pages. £11.65.
CONTENTS: Introduction to audio technology. Audio preamplifiers. Audio power amplifiers. FM tuners and stereo decoders. Electrophonic music systems. Quadrophonic sound systems. Electronic organs. Audio system interference.

Amateur Antenna Tests and Measurements
HARRY D. HOOTON. Howard W. Sams, Indianapolis, 1977. 21 × 13·5 cm. 173 pages. £6.95.

Radio Direction Finding and the Resolution of Multicomponent Wave-fields
P. J. D. GETHING (*Admiralty Surface Weapons Establishment*). Peter Peregrinus, Stevenage, 1978. IEE Electromagnetic Waves Series no. 4. 14 × 22 cm. 319 pages. £13.50.
CONTENTS: Ionospheric modes. Wave-field models. Zero-aperture bearings in two-ray wave-fields. Directive array patterns. Instrumental and site errors. An introduction to resolution techniques. Wave interference effects for circular arrays. Wave interference effects for interferometers. Wavefront analysis: the concept. Wavefront analysis using imperfect data. Ray paths. The effects of ionospheric tilts. Bearing accuracy and d.f. plots.

Frequency and Time
P. KARTASCHOFF (*Telecommunications Department, Swiss Post Office*). Academic Press, London. 1978. 15 × 24 cm. 218 pages. £12.50.
CONTENTS: Frequency stability measures. Standard frequency generators and clocks. Time scales. Frequency and period measurements by means of counters. Phase-time measurements. Frequency domain measurement techniques. Radio signal comparison methods.

Elektronika Linearna Integrisana Kola (Electronic Linear Integrated Circuits) (In Serbo-Croat).
BRANKO RAKOVIC (*Faculty of Electrical Engineering, University of Belgrade*). Gradevinska Knjiga, Belgrade, 1979. 24 × 17 cm. 365 pages.
CONTENTS: High frequency characteristics of bipolar transistors. Field effect transistors at high frequencies. Multistage amplifiers. Feedback amplifiers. Amplitude and phase responses of feedback amplifiers. Compensation in operational amplifiers. Integrated operational amplifiers. Wide band amplifiers. Noise.

Introduction to Microelectronics
2nd Edition
D. RODDY (*Professor of Electrical Engineering, Lakehead University*). Pergamon Press, Oxford, 1978. 21 × 14·5 cm. 211 pages. £10.00 hardcover. £5 limp cover.
CONTENTS: Basic theory. Processing of silicon devices and circuits. Silicon planar devices and integrated circuits. Bipolar logic circuits. Differential amplifiers. Metal-insulator semiconductor devices. MOS circuits. Thin-film circuits. Thick-film circuits. Hybrid circuits. Microwave applications of microelectronics. Semiconductor memories.

Digital Processing of Speech Signals
L. R. RABINER (*Acoustics Research Laboratories, Bell Telephone Laboratories*) and R. W. SCHAFER (*School of Electrical Engineering, Georgia Institute of Technology*). Prentice-Hall, Englewood Cliffs, N.J. 1978. 18 × 24 cm. 508 pages. £17.50.
CONTENTS: Fundamentals of digital speech processing. Digital models for the speech signal. Time-domain models for speech processing. Digital representations of the speech waveform. Short-time Fourier analysis. Homomorphic speech processing. Digital Linear predictive coding of speech. Digital speech processing for man-machine communications by voice.

The Architecture of Small Computer Systems
A. G. LIPPIATT (*School of Information Sciences, Hatfield Polytechnic*). Prentice-Hall International, Hemel Hempstead, 1978. 21 × 14 cm. 161 pages. £5.95 (hard back), £3.95 (paperback).
CONTENTS: Introduction to a computer system. The instruction set of the central processor. The coding of information in computer systems. Addressing the memory. Data transfers withih the computer system. Interrupts. Computer arithmetic.

Man-Made Radio Noise
E. N. SKOMAL (*Aserospace Corporation, El Segundo*). Van Nostrand Reinhold, New York, 1978. 23·5 × 16 cm. 332 pages. £16.15.
CONTENTS: Definition, distribution and sources of radio noise. Automotive noise. Electric-power generation and transmission line noise. Industrial, scientific, medical and transportation sources of radio noise. Theory of the envelope statistics of man-made radio noise. Composite, metropolitan-area, surface man-made noise. Composite, metropolitan-area, elevated and airborne incidental noise.

The Microprocessor and its Application
Edited by PROFESSOR D. ASPINALL (*Formerly University College of Swansea*). Cambridge University Press, 1978. 23·5 × 15·5 cm. 357 pages. £12.50.
CONTENTS: Hardware Components: Aspects of integrated circuit hardware. Processing elements—PMS. Processing elements—the instruction set processor. Processing elements—Input, output and concurrency. Software Techniques: The hardware-software spectrum. Introduction to system design. Logical design. Design simulation and optimization. Design implementation. High level languages. Correctness in dedicated systems. Implementation of high level programs. Aids to program maintenance. Management of Microprocessor Application Projects: Microprocessor project management. Hardware/software tools for microprocessors.

John Logie Baird and Television
MICHAEL HALLETT. Priory Press, Hove, 1978. 24 × 20 cm. 90 pages. £3.95.

Electrical Principles for Technicians 2
S. A. KNIGHT (*Bedford College of Higher Education*). Newnes-Butterworths, London, 1978. 24·5 × 18·5 cm. 119 pages.
CONTENTS: Units and definitions. Series and parallel circuits. Electrical networks. Capacitors and capacitance. Capacitors in circuit. Magnetism and magnetization. Electromagnetic induction. Alternating voltages and currents. Magnetic circuits. Reactance and impedance. Power and resonance. A.c. to d.c. conversion. Instruments and measurements. Alternating current measurements.

Generation of Precision Artwork for Printed Circuit Boards
P. LUND (*Radiometer, Copenhagen*). John Wiley & Sons, London, 1978. 23·5 × 15·5 cm. 337 pages. £13.50.
CONTENTS: Base materials. Board manufacture. Purchase specification. Layout of the board. Generation of precision artwork. Theory. Contact printing on diazo film. Preparation of precision artwork. Practice. Automated generation of artwork/filmwork. Design of printed circuit boards for automatic component insertion.

New and Revised British Standards

Copies of British Standards may be obtained from BSI Sales Department, 101 Pentonville Road, London N1 9ND.

RELIABILITY DATA ON ELECTRONIC COMPONENTS

A revised edition of BS 4200 **Guide on the reliability of electronic equipment and parts used therein:** Part 3 **Presentation of reliability data on electronic components (or parts)** (£6.40) has recently been issued by BSI. The increasing complexity of electronic equipment, the very large number of components used and the requirements for reliability present serious difficulties to the circuit and equipment designer. Information on ratings, characteristics, dimensions, etc., is normally supplied in specifications, catalogues and data sheets. Nevertheless, to be able to predict the operating performance of equipment, detailed data are necessary on the changes in characteristics and on the failure rates of the components.

BS 4200 Part 3 is intended to provide guidance for presenting data necessary to distinguish the reliability characteristics of a component. The data may be those relating to failures and failure rates, or relating to changes (or drift) of characteristics. Such factual information should be available to the circuit and equipment designer to enable him to assess correctly the reliability of his circuits and units. This information will be obtained from reliability tests made on the electronic components in laboratories, and should be presented as indicated in this standard.

Requirements are in four sections; identification of components tested, test conditions, data on failures and data on changes in characteristics.

CATHODE-RAY OSCILLOSCOPES

BSI has revised BS 4739 which is issued under the new title **Expression of the properties of cathode-ray oscilloscopes** and is identical with IEC Publication 351. It is published in two parts which give rules for standardizing methods of expression and define appropriate terminology and catalogue data.

Part 1 **General** (£12.50) deals with the following types of cathode-ray oscilloscopes:

General purpose, for measuring electrical quantities
Multi-trace
Assemblies with detachable or incorporated parts (e.g. probes or interchangeable plug-in units)

The qualities of cathode-ray tubes are considered only when necessary for the evaluation of oscilloscopes.

The companion document, Part 2 **Storage oscilloscopes** (£4.50), applies to oscilloscopes incorporating special devices in the cathode-ray tube which permit a signal to be recorded without recourse to the ability of the screen coating to retain the trace for a limited period. It does not apply to oscilloscopes having a normal tube with storage devices external to the tube. Certain types of storage oscilloscopes covered by Part 2 are also capable of operation without storage and, when so operated, comply with the requirements of Part 1.

Both Parts give conditions and methods for testing so that manufacturers' claims of compliance may be verified. Safety requirements, however, are not specified since these are dealt with separately in the revised BS 4743.

SAFETY OF ELECTRONIC MEASURING APPARATUS

The British Standards Institution has just issued BS 4743 **Safety requirements for electronic measuring apparatus** (IEC 348: 1978) (£12.50). This standard results from work done in the International Electrotechnical Commission and is of particular interest to manufacturers of electronic measuring apparatus and users, such as laboratories. It is a revision of BS 4743 published in 1971.

The standard covers definitions, general requirements, the conduct of tests, reference conditions for test purposes, testing under fault conditions, marking, hazards from emanations (ionizing radiation, microwave radiation, laser radiation, poisonous and injurious gases, ultrasonic pressure), effects of heating (including mechanical strength at elevated temperatures), implosions and explosions, electric shock hazards, mechanical strength, general requirements for components, terminal devices, external cords, and information to be provided for the user. Appendixes deal with safety classification, a list of tests, and IEC publications to which reference is made (cross-reference to equivalent British Standards are also given).

AMPLIFIERS AND INTERFACES FOR AUDIO EQUIPMENT

Two further parts of the British Standard BS 5428 **Methods for specifying and measuring the characteristics of sound systems equipment** have recently appeared.

Part 2 **Amplifiers** (£6.40) is in two main sections, one giving general requirements, the other characteristics to be specified (and the relevant measuring methods). The characteristics include gain, response, amplitude non-linearity, noise, hum, linearity unbalance, external influences, magnetic stray field, cross talk and separation, and alignment. It also provides classification for the characteristics to be specified.

Part 10 **Preferred matching values for the interconnection of sound system components** (£7.50) gives the preferred electrical values for the correct interconnection of sound system components. In the new standard, one part of a system is said to permit correct electrical interconnection to another part when they are so designed as to give satisfactory operation when connected together. In this context, 'correct interconnection' does not necessarily imply maximum power transfer between units. It covers equipment for broadcast and comparable use and also equipment for public address and domestic use. It also provides graphical interconnection symbols for use on equipment.

Other parts of BS 5428 already published are Part 1 **General,** Part 3 **Microphones,** Part 4 **Auxiliary passive elements,** Part 6 **Headphones,** Part 7 **Automatic gain control devices,** Part 8 **Artificial reverberation, time delay and frequency shift equipment** and Part 11 **Loudspeakers.**

TWO NEW STANDARDS FOR PROXIMITY SWITCHES

The series of British Standards for proximity switches for use in production equipment, mechanical handling, packaging and general application has recently been extended with the publication of two new standards.

BS 5683 **Low voltage switchgear and controlgear for industrial use. Inductive proximity switches. Form C, for direct current 3 or 4 terminals** (£4.50) specifies two types of switches, rectangular in form and having a square section, designed for use with a rated supply voltage not exceeding 48 V d.c. The types referred to are C 21·2, having a front sensing face, and type C 21·2, having an upper sensing face, both of which are depicted by means of diagrams.

The second new standard is BS 5684 **Low voltage switchgear and controlgear for industrial use. Inductive proximity switches. Form D, for direct current 3 or 4 terminals** (£4.50) which specifies a switch having a rectangular form with a rectangular cross section, designed for use with a rated supply voltage not exceeding 48 V d.c., and is depicted by means of diagrams.

Maximum sizes, mounting dimensions and a designation system are specified. Also included are various performance requirements such as the operating distance, the repeat accuracy, the differential travel and the operating frequency. Both these standards, in common with the rest of the series, are British versions of equivalent European documents which are being adopted by the other EEC countries as national standards.

THE SYSTEMATIC APPROACH TO QUALITY

A British Standard which, if applied, could help provide Britain with a new reputation for quality in industry has just been published. It draws upon the largely untapped experience of the country's most determined and effective users of quality management systems. The Standard does not make reference to any specific product or industry; 7000 other standards do that. It presents practical requirements of quality assurance which are absolutely fundamental, in a form so concise that it may be used as the yardstick against which a firm's capability of undertaking a contract is assessed. 'Quality', for an organization working to BS 5750 **Quality systems** will not mean just the assiduous correction of defects and failures, but a practical discipline which integrates objectives throughout the industrial process, from design to marketing and servicing.

BSI has been tackling quality systems for some time by formalizing terminology and recommending systematic procedures wherever possible. Guides to the operation and evaluation of assurance systems based on the comprehensive quality control requirements of NATO and the Ministry of Defence were published in 1972 and 1974. Experience gained by industry in applying these recommendations has enabled them to be transformed into requirements, and BS 5750 is the result. It is a turning point in the field of standards, providing the word 'quality' with a new authority which allows it to be used objectively in contracts and evaluation reports. What is more, for all its far-reaching scope, BS 5750 is succinct, presenting the requirements in three easily assimilated documents.

BS 5750 specifies requirements for three basic levels of system for the assurance of quality of products or services, or both. These are given in three parts of the standard. They cover subjects such as the organization, review and planning of the quality system; the necessary control of various functions and products, the records and documentation required to demonstrate that controls are effective and the training that may be needed to operate the system effectively.

Part 1 **Specification for design, manufacture and installation** (£2.60) deals with the quality system to be applied when the technical requirements are specified principally by performance or where the design has not been established. Part 2 **Specification for manufacture and installation** (£2.60) is concerned with the system to be applied when requirements are specified in terms of established design and manufacturing practice, but where conformity can be ensured only by inspection and test during manufacture and installation. Part 3 **Specification for final inspection and test** (£2.60) specifies the quality system to be applied when conformity with requirements can be adequately established by inspection and tests on finished products or services. The new standard can be used in a number of ways. It may form the basis for evaluating a supplier's quality management system or be invoked in a contract to specify appropriate quality assurance requirements. It may also be used in other documents, such as product standards, where reference to a quality management system is appropriate.

COMPONENTS FOR ELECTRONIC EQUIPMENT BASIC TESTING

A number of new parts of BS 5772 **Electromechanical components for electronic equipment basic testing procedures and measuring methods** have just been issued by BSI. They are:

- Part 2 · General examination electrical continuity and contact resistance tests, insulation tests and voltage stress tests (£5.50).
- Part 3 Current carrying capacity tests (£4.50).
- Part 4 Dynamic stress tests. (£4.50).
- Part 5 Impact tests (free components), static load tests (fixed components), endurance tests and overload tests (switches) (£5.50).
- Part 6 Climatic tests and soldering tests (£7.50).
- Part 8 Connector tests (mechanical) and mechanical tests on contacts and terminations (£4.50).
- Part 9 Cable clamping tests, explosion hazard tests, chemical resistance tests, fire hazard tests, r.f. resistance tests, capacitance tests, shielding and filtering tests and magnetic interference tests (£4.50).

These basic specifications together with BS 5772 Part 1 **General** and BS 5772 Part 7 **Mechanical operating tests and sealing tests,** form a nine-part standard defining basic testing procedures and measuring methods for the following families of electromechanical components:

Solderless connections
Connectors for frequencies below 3 MHz
Sockets for electronic tubes
Sockets for other plug-in devices
Lever switches
Push-button switches
Rotary switches
Sensitive switches
Thermal time delay switches
Thermostatic switches

The standard contains fundamental information on test methods and procedures and is intended to be used in those cases where a generic specification for a certain component has been prepared, so as to achieve uniformity and reproducibility in the testing procedures.

The requirements for the performance of the components are not covered by this standard. The relevant specification for the item under test defines the permissible performance limits.

These parts of the standard are identical with IEC Publication 512 Parts 2 to 9.

STANDARDIZATION IN PRINTED CIRCUITS

Two new British Standard specifications related to printed circuits: one an international specification for a grid system, and the other part of the BS 9000 system for electronic components of assessed quality.

BS 5830 **Grid system for printed circuits** (IEC 97) (£2.60) defines a grid system to ensure compatibility between the printed circuits and components to be mounted on them at the intersections of the grid. The original system based on a 2·54 mm spacing is retained and the need of modern techniques for smaller spacings is satisfied by means of a grid system based on a 0·1 mm nominal spacing.

BS 9764 **Printed circuits without through hole connections** (£7.50) specifies tests taken from those listed in a generic document (BS 9760), standard test patterns and the requirements necessary for obtaining capability approval. A test schedule for quality conformance inspection during manufacture is given together with an example of a customer's detail specification.

BBC Digital 4-Field Standards Converter Enters Service

A new BBC-designed digital television field store standards converter has entered service at the Television Centre where it is used to convert pictures between the 525-line 60-field NTSC system used in many parts of the world, including the USA and Japan, and the 625-line 50-field PAL system used in the United Kingdom and much of Europe. The converter is equally good at conversion in both directions and is being used extensively in 625/525 mode to 'export' BBC programmes, as well as in the reception of news and other programmes from overseas. The new converter replaces the world's first electronic field store standards converter, the FS-series equipment which has been in service at the Television Centre since 1968. Like the FS equipment, the new converter also makes a breakthrough: it is the first operational standards converter which stores four television fields (two complete pictures). Studies have shown that converters storing only two fields (one picture) suffer from certain technical limitations on their conversion performance.

In designing the converter, advantage has been taken of developments in semiconductor technology to achieve improved performance compared not only with the FS analogue equipment but also with later digital designs. In particular the availability of dynamic m.o.s. random access memories has led to the provision of cheaper, more compact field stores. The configuration permits easier access to information than could be achieved with earlier semiconductor storage devices. The introduction of specially-designed comb filters in the input decoders to separate the chrominance signal from the luminance signal prior to the standards conversion process has also contributed to improved overall quality. The converter is shown diagrammatically in Fig. 1.

Though its primary function is as a standards converter, the new equipment is required to provide the signal processing for all international programme exchange. It can be used as a high-quality transcoder to change SECAM signals to PAL and vice versa, and also as a synchronizer for correcting the timing of one television signal relative to another or for 'repairing' an input which is out of specification.

Through the use of external decoders the converter can be used with the entire range of internationally-recognized systems including PAL M (525-line PAL). A feature is that where the incoming programme—news items from different sources for example—contains different standards, the equipment automatically switches to the correct mode.

Another important feature of the new converter is that the design includes a microprocessor-based monitor unit to maintain surveillance of power supplies, bay temperatures and input/output conditions. The monitor not only gives an indication (in plain language on a video display screen) of a fault condition but also classifies the grade of remedial action required. Digital test routines are also provided so that the operation can be checked when the equipment is not being used for programme conversion.

Background to the new converter

The task of converting a colour picture originating on the 525-line 60-field NTSC system to the 625-line 50-field PAL or SECAM system (or vice versa) involves three major problems: changing the number of fields per second; changing the number of lines, and changing the colour television system. Of these, changing the number of fields is very much the most difficult and until the introduction of the BBC's FS-series equipment in 1968 the conversion could only be achieved using a camera operating on the required standard to view a television picture display of the incoming signal. The FS-series equipment performed the field conversion by using quartz delay lines to store the incoming fields so that their picture information could be redistributed to field periods corresponding to the outgoing standard. The store size was one field period.

When the FS-series equipment came into service it could probably be described as being ahead of its time and its facilities have been used by broadcasting authorities all over the world: however, in 1976 it was decided that new equipment, with superior performance, would be required for programme interchange during the 1980's. Commercially available converters were examined but none reached the necessary standard of performance. In anticipation the BBC Research Department had undertaken extensive analysis of the fundamental processes involved in field standards conversion and were able to show that the interpolation process could be defined scientifically and the parameters mathematically deduced. They recommended that the converter should interpolate from four incoming television fields in order to provide an output with virtually no visible impairments. Trials confirmed the validity of this recommendation and in 1977 the BBC Designs Department drew up a specification for a new 4-field converter and undertook the design and production of

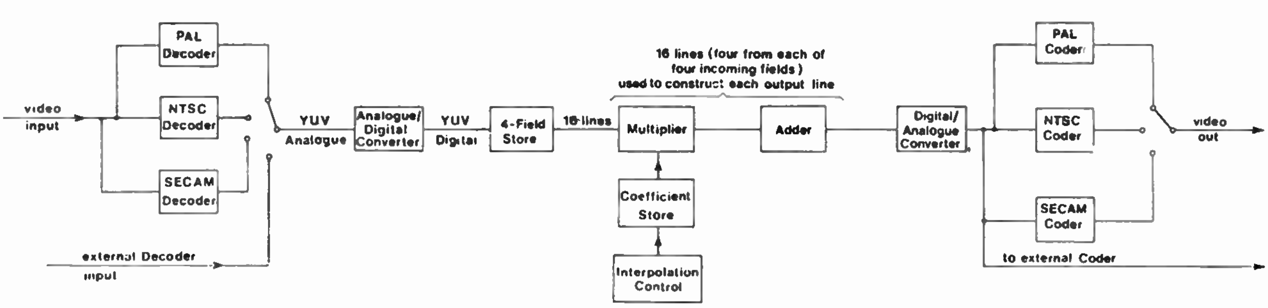


Fig. 1. Functional block schematic of the new BBC digital field store standards converter

two equipments. The first converter, comprising two 7 ft. bays of equipment, was completed last July and, following testing, was brought into operational use in September.

The new converter has already been used for tape-to-tape 525/60 to 625/50 conversions as well as live transmissions from the USA on a regular basis for news inserts. Conversions from 625/50 to 525/60 have included tape-to-tape transfers and direct transmissions to the USA via satellite. Assessment of the quality of the conversions from the new equipment has confirmed that the design objectives have been met: there is now virtually no detectable loss of picture quality in changing the transmission standards of a programme.

Some technical features of the new BBC field store standards converter

Interpolation. To change from one television field standard to another without noticeable impairment of the output picture quality, each output line must be synthesized from several lines of the input signal. This process is known as interpolation. Interpolation from lines within any one field is known as vertical or spatial interpolation, while interpolation using lines from different fields is known as temporal interpolation. Most digital field store converters store two fields to provide a two-field 'aperture function', but it has been shown that the use of an aperture function four fields wide can provide a much closer approximation to the characteristics required for good movement portrayal. By adopting this wider temporal aperture and combining the temporal and vertical interpolation processes instead of handling them separately, as in previous designs, the conversion quality of the new converter approaches closely to the optimum.

In operation, the digital store is used to hold four fields of the input signal. The interpolator has access to four consecutive lines in each of these fields simultaneously, i.e. a total of sixteen lines are used in a single interpolation process to synthesize a line of signal at the output standards. The relative positions of the set of input lines are shown in Fig. 2. The actual contribution that an input line makes to the output line depends upon its temporal and vertical positions relative to the required output line and is controlled by an appropriate multiplier coefficient held in a programmable read-only memory. The process of interpolation then consists of multiplying the sixteen input samples by sixteen coefficients and summing the resultant products. The colour and luminance signals are treated separately and the use of optimum apertures for each results in reduced video noise at the output compared with the input.

Any further developments in the operating characteristics of the converter can be accommodated simply by inserting a p.r.o.m. having the necessary changed interpolation coefficients.

Digital Store. In converting from the 625-line 50 field PAL system to the 525-line 60 field NTSC system the input signal is sampled at 15·94 MHz; in the reverse direction the sampling rate is 15·73 MHz. The samples, coded to eight-bit accuracy and handled in parallel form, are stored with luminance Y and colour difference samples (UV) interleaved in the five-word sequence $YUYVY$. Successive lines from one field are fed in sequence into each of four store blocks, i.e. for a given field, block A will hold lines 1, 5, 9 . . . block B will hold lines 2, 6, 10 . . . and so on. The basic storage element used is a 4k bit dynamic memory which is an industry standard.

Decoder. Conventional decoding techniques provide a luminance signal with excessive residual subcarrier, noticeable distortion through inadequately filtered chrominance sidebands, and a severe dip in amplitude response which attenuates fine picture detail. The PAL and NTSC decoders employed in the new equipment use comb filters which considerably reduce these imperfections and make a substantial contribution to the excellent performance of the converter.

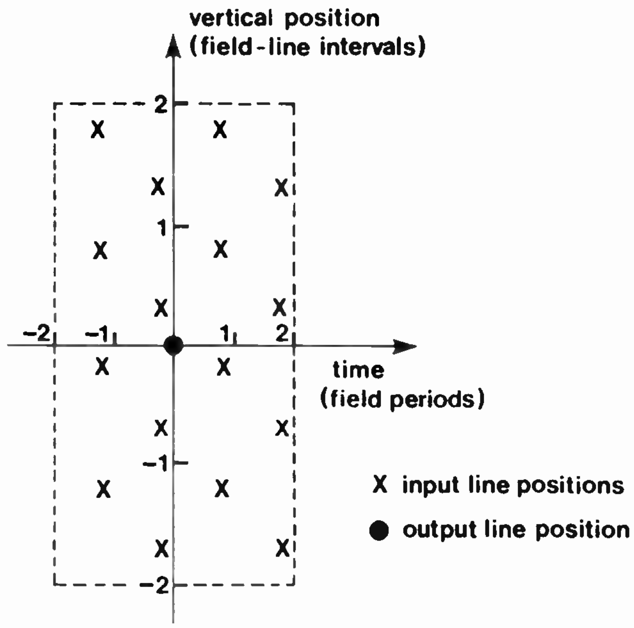


Fig. 2. Positions of sixteen input standard lines used to produce one line of the output standard

UDC 621.391.7: 621.395/6
Indexing Terms: Speech secrecy systems

The development of speech encipherment

N. R. F. MacKINNON,
M.B.I.M., C.Eng., F.I.E.R.E.*

SUMMARY

This paper surveys some past and present principles of speech coding. The advent of digital logic systems and integrated circuit components is seen as a major turning point in the implementation of effective coding principles that could not previously be used satisfactorily.

Various coding methods are discussed, covering frequency domain, time domain and analogue and digital output systems. The problems of key generator integrity, synchronization, received voice quality and recognizability, and channel bandwidth restrictions, and the security offered by different systems are considered.

The author concludes that in the immediate future high security speech encipherment systems will be of the digital output crypto-vocoder type, with audio bandwidth analogue output coders offering an increasingly secure alternative for the majority of applications. Digital output coders will continue to be used effectively in those applications where the communication equipment and transmission path have the necessary characteristics to carry low-error-rate digital signals. Advances in analogue-to-digital speech conversion providing acceptable quality at bit rates below 9·6 k bit/s should eventually enable a simple digital output coder to be as effective as a crypto-vocoder for audio band high security applications.

* *Consultant, Spex Hall, South Newton, Salisbury, Wiltshire.*

1 Introduction

1.1 Voice Coding Prior to 1945

It is almost 100 years since the first voice coding system was patented in America, only five years after the introduction of the Bell telephone in 1876.[1] The coding system was simple: sequential segments of speech were transmitted over different circuits, making it necessary for the interceptor to monitor simultaneously as many as ten lines to receive useable intelligence. It can be argued that such a system is not in truth a speech scrambler as none of the parameters of speech such as frequency, amplitude, pitch or sequence are disturbed.

It was not until after the 1914–1918 World War that true scramblers using simple frequency inversion techniques became available. The Bell Laboratories refined frequency domain (f.d.) techniques and by the late 1930s had produced band-splitter inverters that were in widespread use by the US Government and telephone companies. The San Francisco–Honolulu telephone circuits were protected by such devices as early as 1937, superseding the inverters previously used.[2] It is of interest that the US Government did not use this protected speech link on 6th December 1941 to warn of the probable attack on Pearl Harbour as they suspected, rightly, that the equipment was not immune to Japanese intercept and analysis; the warning message was sent in a textual cipher by commercial circuits, being delivered after the successful attack on the 7th.

It is known that band-scramblers were also developed and in diplomatic use by the late 1930s. Their basic coding principle was the transposition of four or more frequency bands within the audio spectrum, some inverted. Band-scramblers were extensively used in the 1939–45 war. In spite of their increasing sophistication both sides developed equipment that could decode band-scrambled transmissions in near real time. The success of the German 'Postal Research Bureau' in decoding the band-scrambled UK–US radio telephone circuit has been well documented. On two occasions significant conversations between Churchill and Roosevelt were successfully intercepted and decoded.[2]

1.2 The Advent of Semiconductor Circuit Techniques

Though the theoretical possibilities of other voice coding techniques were well understood it was not until the development of digital logic and semiconductor circuits in the 1950s that voice coding methods independent of frequency domain changes became practical propositions. The first significant advances made possible by digital techniques were the evolution of pseudo-random digital key stream generators and effective digital synchronization circuits that made the use of fast changing code patterns possible. Voice ciphers, analogous in some ways to the well-established electro-

0033–7722/80/040147+09 $1.50/0

mechanical textual coders in which a pseudo-random key stream of 0's and 1's was added to the Murray or Baudot 0's and 1's representing the clear text, became practical equipments. The principles of a digital voice cipher are the same, the Murray or Baudot code 0's and 1's being replaced by a digital stream derived from level sampling or quantization of voice, this output being then added to the key stream as in the textual ciphers, albeit at a rate of 18 k bit/s rather than 60 bit/s. The resultant coded output sounds like white noise to the interceptor.

The 50 kHz channel bandwidth requirement of such a signal confined the use of digital output coders to v.h.f. and u.h.f. radio circuits, and was in part responsible for the military preference for v.h.f. battle radios by Western Governments in the 1960s and 70s. H.f. s.s.b. transceivers, superior in many aspects to their v.h.f. counterparts, could not provide the channel bandwidth required by digital output coders, or the transmission path frequency stability necessary for effective use of frequency domain scramblers. 'White noise' coders gained rapid acceptance in Middle East and S.E. Asian tactical v.h.f. battlefield environments and continue to give secure service today.

Another major development made practicable by semiconductors and i.c. logic was time element scrambling (t.e.s.), often also referred to as t.d.m.[3] In this system segments of digitized speech are stored in shift registers and then extracted in a coded non-sequentially delayed order. A digital-to-analogue conversion takes place prior to transmission, allowing the t.e.s. analogue signal to be transmitted over any audio channel, including the increasingly used h.f. s.s.b. circuits. Prior to digital logic circuits this system was impractical as analogue delay stores such as tapes or wire were mechanically inefficient and difficult to synchronize. After a hesitant start in the early 70s t.e.s. or t.d.m. coders now appear in the communication security armoury of over thirty nations, primarily because of their good voice quality and flexibility in operational use on s.s.b., telephone and satellite circuits.

1.3 Integrated Circuit Contributions

Many other voice coding techniques have made effective use of i.c. technology. Masking noise coders, in which a varying[4] masking signal was superimposed on the clear speech, became available in the late 1960s. A variation of this technique, in which an echo signal is used as the mask, is still available as an effective coder. Other variations were frequency or pitch distortion, achieved by varying the speech quantization rate at the input and its digital-to-analogue conversion at the output. The effect is the same as randomly-timed frequent speed changes on a tape recorder. Another technique is to store digitized speech for say, 100 milliseconds, and then to extract it backwards. Most of these methods offer a low level of security in themselves, but in microprocessor-controlled combinations they can be a formidable barrier against analysis.

In the last five years considerable advances have been made in voice cryptophony by the use of vocoders, in which the non-redundant elements of speech are analysed, digitized and transmitted. At the receiver speech is artificially synthesized. Vocoders can pass intelligible speech at bit rates as low as 1200 bit/s and at 2400 bit/s speaker recognition is possible. A vocoder is in itself an expensive single code voice cipher. If the bit stream is enciphered by an ordinary digital cipher equipment the resultant crypto-vocoder is as secure as the cipher equipment's key generator (k.g.) permits. Crypto-vocoders are being increasingly used on the more sensitive Government, diplomatic and military circuits. They will work effectively over most audio channels, albeit at the expense of voice quality.

In retrospect it can be said that the techniques of speech coding developed slower than the techniques of intercept and analysis.[5] Recently the prevalence of industrial espionage, allied to continuing worldwide diplomatic and military confrontations, have accelerated advances in the security and use of voice coders. Crypto-level security is available though expensive, whilst at the other end of the scale simple coders costing £100 or so can provide privacy for telephones, police v.h.f. patrol sets and other users requiring only a short security period, or 'cover time' for their transmissions. Many of the more secure equipments now available use combinations of frequency and time domain coding when a narrow band analogue output is required. Digital speech coders using a novel divide-and-multiply feedback shift register coding system have been commercially available since 1977.[6] Frequency hopper/inverters offer yet another technically updated coding system.[7] The advent of charge-coupled devices (c.c.d.s) is opening up new possibilities in signal analysis and modulation. Necessity is spurring designers to use emerging microprocessor technology to the full, and a voice coding history of the present time will show how leisurely was the pace of development in the first fifty years compared with the great developments in the last decade.

2 Voice Characteristics—Scramblers and Voice Ciphers

For coding purposes the human voice is characterized by varying relationships of frequency, amplitude and timing within a speech segment. The distinctive elements of speech are called phonemes, and there is a complex relationship between a phonemic sequence and the acoustic output, varying from speaker to speaker. Voice coding principles can be somewhat arbitrarily but usefully divided into two families—scrambling and ciphering. In scramblers the phonemic sequences are coded in the frequency or time domains, or both. The ciphering process is to digitize the clear voice signal and then modify the values of the digits by interaction with a digital keystream. This ciphering process destroys the transmitted voice characteristics, whereas the scrambling process distorts them. However, voice characteristics have a bearing on both scramblers and voice ciphers, as the latter require a high output bit rate if they are to preserve speaker recognition and voice quality, both priority requirements of users.

Speech is rich in redundancies. The coding of these apparently valueless redundancies, coupled with the pitch and syllabic rhythm of the speakers voice are the main factors in permitting easy speaker recognition and intelligibility. Coding processes, such as low bit-rate output vocoders or digital output coders, that reduce redundancy also reduce speaker recognition and audio quality, and are unpopular with users. The robot-like output of some vocoders is an example. Though the intelligence is present and understandable, the partial elimination of redundancies and incorrect pitch assessments mask the emotions and personality of the speaker to an unacceptable degree.

Thus a voice message is made up of two sets of intelligence. One is the content of the message, the other is the identity, manner and personality of the speaker. Ideally a coded voice transmission should shield both message content and speaker characteristics from the interceptor, and reproduce the clear speech faithfully to the authorized receiver. Scramblers and voice ciphers in their simpler less expensive forms tend to do one or the other well, but seldom both. Good voice quality and speaker recognition are typical attributes of scramblers, whilst high integral security levels are more usually associated with digital output voice ciphers.

In the following Section of this paper various methods of scrambling and voice ciphering are considered. Many of their design features are influenced by the necessity to encode and reproduce voice characteristics as well as intelligence. This may well explain the continuing development of time-domain-based scramblers, even though the security integrity of voice ciphers is generically better.

3 Coding Methods—Scramblers

3.1 Frequency Domain

The basis of many f.d. scramblers is the balanced modulator. If a carrier signal frequency of f_1 is modulated by an input frequency f_2 the output will contain the upper and lower sidebands f_1+f_2 and f_1-f_2, the latter being inverted. It can be readily seen that variations of f_1 will result in variations in the frequency limits of suitably filtered output bands, which may now lie outside the pass band of the audio circuit to be used. Another balanced modulator (or modulators in a multiband split equipment) translates these frequencies back into the speech band.

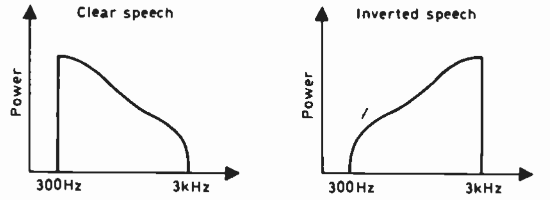


Fig. 1. Frequency inverter in which speech is band limited.

3.2 Speech Inverters

This is the simplest form of f.d. coding, in which the high frequencies are translated to low frequencies and low to high by taking the lower sideband from a balanced modulator with a suitable carrier frequency f_1. In effect this is a single code equipment offering privacy from overhearing only. Figure 1 illustrates this principle.

3.3 Bandshift Inverters

Speech is inverted and that portion of the inverted spectrum that lies above the audio passband is translated and added to the lower end of the passband, as may be seen in Fig. 2. The inversion frequency F can be varied in fixed discrete steps or in a key-generator-controlled pattern. These equipments are called frequency hop inverters.[7] Variations in the value of F and the time interval between changes in F permit frequency hop inverters to offer many coding permutations.

3.4 Band-splitters

Since the 1930s bandscramblers have dominated the frequency domain voice scrambling field. The audio band is filtered into a number of sub-bands, typically five. These sub-bands are translated and possibly inverted to other slots in the transmitted frequency spectrum. Figure 3 illustrates the principle. There are $5! \times 2^5$ (3840) possible fixed combinations in a five-band system, but only a dozen or so provide good mutual security between codes and some security against simple band filtering analysis. However, if the translation frequencies for the sub-bands and the timing intervals between changes are pseudo-randomly-controlled by the

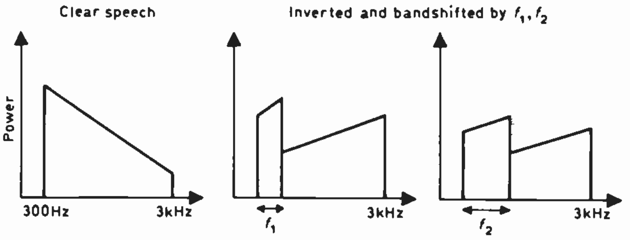


Fig. 2. Bandshift and frequency hop inverter. After inversion the spectrum is shifted along the frequency axis by between 50 and 350 Hz. Inverted speech outside the upper bandlimits is cut off and added at the bottom. Changes in F, the inversion frequency, provides the coding variables.

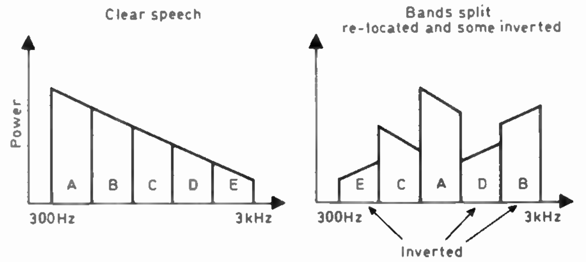


Fig. 3. Bandsplitters. A pseudo-randomly changing pattern of band relocation and inversion greatly increases the number of valid codes.

key generator a large number of code patterns are available to the user. The major weakness of bandsplitters lies in the energy/frequency relationships in human speech. Typically 40% of the energy will lie in the 400–800 Hz area, and this energy content will be present in the new frequency translated band, say 1800–2400 Hz at the output. This helps the analyst to pick off the A or high energy band in each time frame and deduce the base keys and k.g.-imposed pattern.

3.5 Time Element Scramblers

T.d.m. coders, as they are usually called, rely on the partial inability of the human brain to interpret non-sequentially received speech segments. Figure 4 shows the principle of time division coding, and Fig. 5 is the block diagram of a typical t.d.m. coder. The output from the eight-segment store can be in a fixed base code, or pseudo-randomly changed by the k.g. selecting other base codes stored in the r.o.m. The number of base code sequences that can be obtained is dependent on the number of base codes stored in the r.o.m. and the k.g. performance parameters. It is usual to find well over 10^{10} different mutually secure sequences in a good t.d.m. coder. This method of coding is sometimes used in combination with f.d. scramblers, which increases the security of the system by an order of magnitude at the expense of complexity and speech quality.

Important factors in design are the duration of individual speech segments, typically 20 to 60 ms, the number of segments in the time frame for each coding pattern, and the coding pattern itself. The storage of speech segments and their transmission in a rearranged order imposes a circuit delay equal to twice the number of segments in a code pattern or frame. Thus a system using eight 30 ms segments in a frame would have an overall circuit delay of $2 \times 8 \times 30 = 480$ ms. Such delays approximate to those experienced on satellite phone calls and are not liked by users. As the security integrity of a t.d.m. coder is improved by a greater number of segments in a frame, and segment duration should be not less than 20 ms for protection against hear-through, the correct balance between the security offered and user acceptability of delays is an important design point.

The human brain has a remarkable capacity for decoding scrambled speech. There are those who can understand inverted speech, others who can extract intelligence from filtered speech samples covering only 300 Hz of the audio spectrum. In a typical t.d.m. system with eight segments per 240 ms frame, the occurrence of a repetitive segment sequence such as 21436587 will probably be understood. The problem of selecting the basic code patterns for a t.d.m. coder is arduous, and though there are 8! ways of rearranging an eight-segment frame not more than 3000 are effective, and many of these are mutually insecure in that the wrong code used at the receiver will produce useable intelligence.

There are several variations on the basic t.d.m. system described. One method, popularly known as the 'sliding window', pseudo-randomly selects stored speech segments from, say, eleven rather than eight stores and passes them to the output. There are no fixed base codes, and the security of the system is dependent on the efficient programming of the key generator to avoid sequences and to provide some measure of mutual inter code security.

Like all other scramblers with an analogue output, t.d.m. coders may be susceptible to mathematical attacks using Fourier series analysis to piece together consecutive segments, thus leading to a better understanding of the underlying k.g. sequences. In order to increase the time taken to analyse and decode intercepted messages various secondary codes can be superimposed on the basic system. Segment boundaries are regularly spaced and can frequently be detected by spectrographic analysis, so one design pseudo-randomly alters the length of segments by up to 18% and reduces clear speech 'hear through' by pitch distortion as described in Section 1.3. A further refinement of t.d.m. coding is achieved by splitting the audio signal into high and low frequency bands before scrambling the time elements in each band in a different pattern. It will be appreciated that each individual coding process inevitably reduces voice quality as well as improving the security of the system. T.d.m. coders incorporating pitch distortion, separately coded high and low bands, and

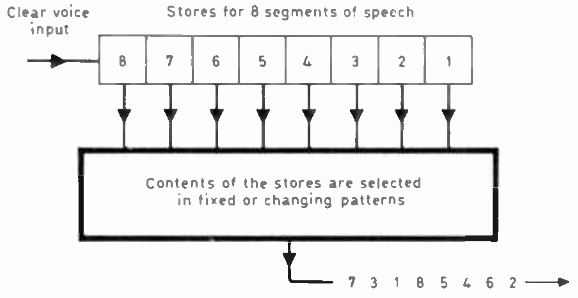


Fig. 4. Time element scrambler principles. The scrambled voice output consists of the original stored speech segments in a delayed non-sequential order.

pseudo-random reversal periods have been made. In addition t.d.m. coders can be allied directly to f.d. coders such as band-scramblers. Such combinations currently offer the most secure analogue output systems available, but require frequency-stable noise-free transmission paths for good audio performance, and suffer the circuit delays inherent in t.d.m. systems.

3.6 Masking Noise Scramblers

This category of coder relies on the summation or multiplication of the plain speech signal with a fixed or varying noise tone, or in some cases with a delayed echo of a frequency sub-band of the plain speech. In its simplest form a noise tone is superimposed over the speech resulting in unintelligibility. In sophisticated systems the wanted signal is multiplied in a phase shifter by a pseudo-randomly-generated noise tone and the resultant, intermodulation products and all, is passed as audio noise to the output. The inherent disadvantages of masking noise systems are the significant losses in signal-to-noise ratios because the masking signal depletes the wanted signal carrier energy, and the difficulty of removing the noise mask at the receiver. However, where noise multiplicative masking systems are used the scrambled output has lost most of the distinctive syllabic rhythms, pitch and sequences of clear speech and spectrograms reveal little to the analyst. To a lesser extent the echo-masking technique also reduces vulnerability to quick spectrographic or Fourier analysis. It is thought probable that the analogue signal processing capabilities of charge coupled devices will lead to the design of multiplicative or additive masking noise coders with better characteristics than have been achieved in the past, both in speech quality and inherent security.

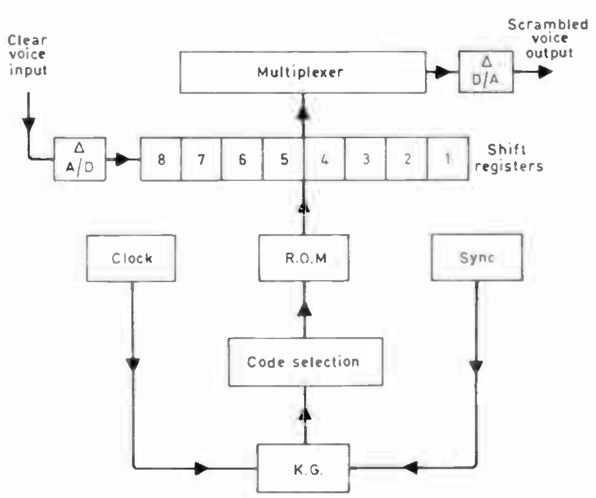


Fig. 5. Time element scrambler block diagram. The k.g. and code selectors determine which pattern in the r.o.m. will be used to multiplex the eight speech segments stored in the shift registers.

3.7 Scrambler Characteristics

Scramblers are characterized by their audio band analogue output, which usually contains the distinctive syllabic rhythms, frequency/power distribution patterns and phonemic sequences of clear speech, albeit in a distorted form. The security integrity of a scrambler depends on the coding method used and its implementation. Fixed coding throughout a transmission is less secure than a coding pattern that changes at short time intervals. If the selection, or generation, of these changing code patterns is controlled by a key generator the security of the equipment is enhanced, and the ultimate security factor becomes the strength of the key generator.

Frequency domain scramblers are susceptible to equipment and transmission path frequency drift. Path noise is a multiplicative rather than an additive factor, particularly in band scramblers. However, f.d. scramblers do not suffer transmission delays and the user requirement for telephone duplex circuits is simpler to implement. Time domain coders, particularly those with a large number of k.g.-selected good codes offer better inherent security than f.d. scramblers, and can suffer frequency drifts of up to ± 40 Hz without loss of intelligence. The circuit delay of up to 800 ms is not liked by users, and makes duplex systems difficult to implement, but this disadvantage of t.d. systems is offset by the better voice quality available.

Scramblers combining f.d., t.d. and masking techniques are complex, but the cover time they can offer is adequate for most applications. Until audio band digital voice ciphers are economically available scramblers will continue to be developed and used, in both governmental and commercial environments.

4 Coding Methods—Voice Ciphers

Voice ciphers are commonly called 'white noise' or 'digital output' coders. They are distinguished from scramblers by the fact that their output carries no overt speech waveform characteristics. The basic philosophy of white noise coders is the interaction of a digital keystream with the digitized clear voice signal. This interaction may be additive, multiplicative, or bit-position shifting. There are two basic systems in use, called in this paper the 'modulo-2' and 'controlled substitution register (c.s.r.)' methods. Another system exists in which the elements of a binary coded number (representing an instantaneous speech power level) are sequence shuffled within the binary word, somewhat analogous to t.d.m. scrambling techniques (Fig. 4). All digital output coders require that the bearer system should have a flat amplitude response from near zero to 8 or 9 kHz, which inhibits their use over lines and radios with passbands of the order of 300–3000 Hz.

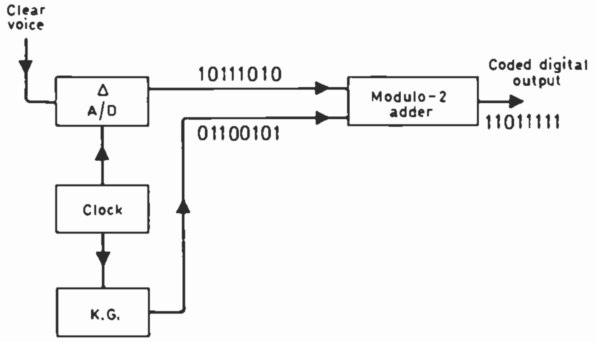


Fig. 6. Modulo-2 voice cipher principles. The pseudo-random digital output of the k.g. is modulo-2 added to the digitized clear voice input.

4.1 Modulo-2

In this system voice is digitized, typically in the delta mode, and then modulo-2 added to a k.g. output.[10] A block diagram of the system is shown in Fig. 6. The security integrity of the system is entirely dependent on the key generator. Recovered speech quality is directly related to the sampling rate of the clear voice input, and the efficiency of the a/d and d/a converters. In practice 12–18 kb/s has proved a useable compromise between acceptable speech quality and channel bandwidth limitations. As coding is on a bit-by-bit basis digit synchronization is vital, and is normally achieved by use of repetitive data bursts during the transmission, sometimes engineered so as to be in the gaps between words or syllables.

Unless precautions are taken raw keystream sequences are transmitted at the start and end of messages, as well as in speech pauses. If a key generator is weak in its non-linearity, cycle length or statistical flatness these sequences of raw keystream can be used to model mathematically the particular key stream in use, and thus to decode the message.

4.2 Controlled Substitution Registers[6]

C.s.r. coders operate by feeding digitized speech, usually derived from a delta modulator, into the feedback (or feedforward) paths of a feedback shift register connected as a multiplier or divider. Figure 7 illustrates the principle. If the switches S remain in fixed positions the digital input is multiplied in a manner determined by the switch state. A development of this basic scheme is to alter the switch positions electronically in a k.g.-controlled sequence, thus making the coding pseudo-random. An n-stage feedback shift register has $2^n - 1$ states, so a 20-stage c.s.r. has over 10^6 fixed codes available. If the feedback conditions, i.e. the switch states, are altered, possibly by using AND gates combining a k.g. input and the fed-back digitized speech to the appropriate modulo-2 adder, the coding is pseudo-random and the number of code sequences obtainable exceeds 10^{6}! In the fixed code case bit synchronization is unnecessary as there are no mandatory starting conditions for the input sequence. If the switch states are changed during a transmission, continuously or at intervals, bit synchronization is required.

C.s.r. coders, whether fixed or dynamically coded, offer good mutual code security. However, they are particularly susceptible to channel noise and consequent error extensions. The effects of such noise can be reduced by using a divider at the transmitter and a multiplier at the receiver, thus spreading the information more effectively in time and limiting transmitted error propagation. In their fixed code form their security depends largely on the number of shift register stages, and hence the number of codes. The fact that bit synchronization is unnecessary is of the greatest value, as all the user required characteristics of delay free, multi-call, duplex or relayed circuits can be implemented without undue technical complexity.

C.s.r. coders can be implemented in two l.s.i. chips, and are currently offered as an integral feature of purpose-designed hand-held transceivers.

4.3 Digital T.E.S.

The coding principle is to rearrange the digitized clear voice signal so that the hierarchical values of adjacent bits or bytes are disturbed in a fixed or pseudo-random pattern. It is in effect a t.e.s. system operating on individual bits rather than speech segments. It is simple to design such a coder, but hard to implement it economically. Only one voice-coding equipment using this system is known to the author, the implementation being effected by use of the IBM DES algorithm for coding digital information.

4.4 Voice Cipher Characteristics

Most current digital output coders are designed as add-on units for specific v.h.f./u.h.f. transceivers. There is a noticeable trend towards designing the transceiver and coder as an integral circuit, permitting reductions in cost, size and circuit complexity. White noise coders operate

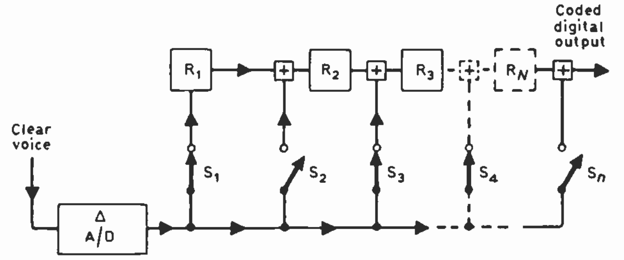


Fig. 7. Controlled substitution register (c.s.r.) voice cipher principles. The position of the switches S determines the feedforward characteristics of the registers R and modulo-2 adders +. This multiplicative process controls the digital substitutions apparent at the coded output.

at between 9 and 18 k bit/s, and until a/d and d/a techniques improve permitting lower operating speeds, they will be restricted in their development. The security integrity of white noise coders can be as strong as that of a textual cipher, though there are pitfalls for the unwary designer. However, properly implemented systems have the potential of offering degrees of security greater than that obtainable from analogue output scramblers.

5 Voice Coding—Crypto-vocoders

It is beyond the scope of this paper to consider the development of vocoders in any detail but their general principles need to be stated.[8,9,10] Vocoders analyse speech and output a low-bit-rate signal, typically 1·2 to 4·8 k bit/s from which they can synthesize the original speech. When the output bit stream is ciphered by, say, modulo-2 addition with the output from a key generator, a crypto-vocoder results. The coded output is suitable for transmission over most audio band circuits, and the security integrity is as good as the k.g. permits. Crypto-vocoders are complex, costly and relatively large, but their operational flexibility and security are such that they are increasingly used in government and commerce.

In the last five years considerable advances have been made in vocoder design, resulting in the availability of crypto-vocoders little larger than a portable typewriter. Two analysis/synthesis methods are in current use. Channel vocoders[8] derive digital control signals from the signal amplitudes in up to 25 frequency channel filters spanning the audio band, and transmit it. At the receiver the original speech is synthesized by a similar bank of channel filters that are excited by random noise or pitch pulses dependent on the incoming control signals. Linear predictive coders (l.p.c.s)[10] operate in an entirely different manner. Speech is defined by determining the characteristic coefficients of a time-varying digital filter that is directly related to the vocal tract output. The coefficients characterizing the digital filter state and pitch information are transmitted. At the receiver these coefficients enable the decoder to predict the required audio output from consideration of previously generated outputs, hence the name linear predictive coding.

Crypto-vocoders are currently the most secure speech ciphers available that can be used over normal telephone circuits. Voice quality is improving but still poor at bit rates of 2·4 k bit/s. Speaker recognition is marginally better in channel vocoders than l.p.c.s. Recent advances in c.c.d. useage and Fast Fourier Transform chips could lead to independent but beneficial advances in vocoder techniques. Currently physiological differences between, for instance, Chinese and European speakers prevent a single vocoder algorithm being effective for both groups. Though costly and complex, crypto-vocoders will, in the author's opinion, become more widely used in the next decade as technological and theoretical advances reduce their cost and improve their recovered speech quality.

6 Synchronization and Key Generators

6.1 Synchronization

Voice coding systems utilizing changing code sequences or timed state changes, such as t.d.m., require synchronization between coder and decoder. In modulo-2 voice ciphers synchronization must be bit by bit, and though the demands of segment synchronization in t.d.m., or band relocation in a band scrambler, are less rigorous, good synchronization is necessary to maintain good quality. In coders controlled by a key generator it is necessary to initiate the generator at the receiver so that it starts at the same point on the key sequence as that of the transmitter. In many of the better scramblers and voice ciphers this start point information is transmitted as part of the synchronization burst. It will be realized that this information, which may itself be coded, only reveals the position on an unknown keystream that has been previously selected by the sender and receiver. Randomizing the k.g. start point for each transmission is good security practice, as it prevents a series of transmissions all being in exactly the same key.

There are two basic synchronization methods used in voice coders. They are commonly referred to as 'start point' or 'PTT sync', and 'continuous sync'. In the former all the necessary timing and key stream start point information is transmitted in a burst sequence at the start of the transmission. In continuous synchronization systems this information is sent in a modulated pilot tone that continuously utilizes 100 Hz or so of the transmitted audio band. Great care is taken in designing synchronization systems so that they are unambiguous and tolerant of noisy and fading circuits.

Start point synchronization systems most commonly use f.s.k. bursts in scramblers and a digit sequence in voice ciphers. The primary requirement is to start the clock, and hence the decoding process, at the receiver. The inclusion of key stream start point information is common, frequently in an automatically generated code related to the preselected key settings. In scramblers it is usual for the synchronization signal to include a preamble tone to stabilize a.l.c. and a.g.c. circuitry prior to the transmission of the tone pattern required to trigger the receiver clock. In digital output systems the synchronization pattern may well contain considerable redundancy or forward error correcting elements to ensure correct reception. Chirp signals are also used for scrambler synchronization as the integration of the signal at the receiver can give an apparent 6 dB gain in the circuit signal/noise ratio.

Start point type synchronization prevents a station which has not received the synchronization burst from joining the communication channel. Late entry, as it is called, is impossible. This is good from the point of view of security and suitable for point-to-point conversations, but unsuitable for multi-station nets of the police or army patrol type. Continuous synchronization systems permit late entry, the penalty being sacrifice of part of the audio band information and consequent degradation of speech quality, as well as an arguable loss of security.

6.2 Revertive Synchronization

Duplex circuits make it necessary for k.g. synchronization systems to take into account the overall transmission delay times between the sender and the receiver. This is effected by revertive synchronization, in which transmission of a synchronization burst causes the automatic generation and transmission of a return synchronization signal by the receiver. This system ensures effective synchronization on duplex circuits, and eliminates synchronization blocking at the start of each transmission, which is particularly undesirable on PTT links.

6.3 Key Generators

The purpose of a key generator is to produce a pseudo-random digital output that can be used to select or generate pseudo-random coding patterns.[11] The ideal k.g. structure has non-linear operative equations with no known general solutions, and a non-repetitive statistically flat output (i.e. similar groups of digits are evenly distributed throughout the key stream). These ideals can not be met easily. The practical measure of k.g. performance is the cost in equipment and time that is necessary to model the mathematical equations governing its operation when given a portion, say a 1000 bits or so, of the output. Key generators in current use are configured from one or more feedback shift registers. By using the output from one generator to interact with another, perhaps by stepping forward, multiplication or periodic addition, key streams of great length can be obtained, thus providing a good code base. A block diagram of a simple k.g. is shown in Fig. 8.[12]

The strength of the key generator is fundamental to the security of digital output voice coders, and is a most important factor in scramblers, where successful audio and visual analysis of parts of the coded signal can often reveal the underlying key stream sequence. Care must be taken that necessary coding inhibits, such as the prevention of transmission of sequential segments in a sliding window t.d.m. system, do not provide an easier entry point to mathematical analysis by the interceptor. A greater number of coding sequence inhibits results in a smaller number of patterns to be identified.

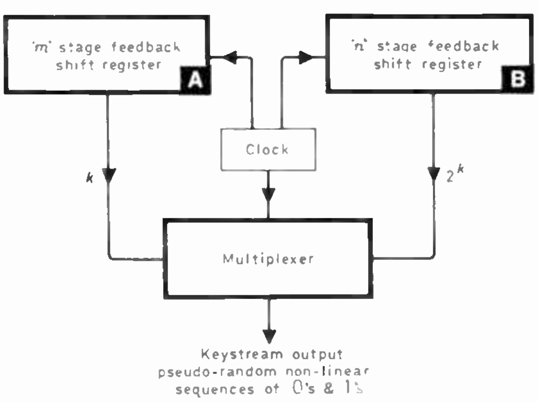


Fig. 8. Keystream generator. The succession of states of A and B have periods of 2^m-1 and 2^n-1 respectively. At any given time the state of A provides a k bit binary address to the multiplexer, which selects one of the 2^k bits of the state of B to be used as the output of the sequence generator. If m and n are coprime, this generator has a non-linear pseudo-random output of period $(2^m-1)(2^n-1)$.

7 Future Developments

User requirements indicate the direction that developments will take. These requirements may appear unrealistic in terms of both theory and practice at a particular time, but become possible as new techniques and components are developed, perhaps for other applications. The current requirements in voice coding are:

A good and quantifiable level of security integrity.
Duplex audio band circuits that operate like ordinary telephones.
Good voice quality.
Minimum circuit delays, not greater than 100 ms.
Ease of operation.

The security integrity of a particular voice coding system cannot currently be measured without detailed knowledge of the resources available to the intercepting analyst. In the future it is likely that k.g.s with improved characteristics will become commercially available, and that analogue implementations of feedback shift registers, perhaps using c.c.d.s in conjunction with analogue output k.g.s, will be developed. The quantification of 'cover time' for any particular system will remain an assessment rather than a measurement, but inherent system security will improve. The development of low-bit-rate voice a/d and d/a systems will eventually lead to voice ciphering systems that can work over ordinary telephone lines and offer acceptable standards of voice quality. In the interim crypto-vocoders will provide better speech quality and become smaller. Linear predictive coders and channel vocoders possess individual characteristics that could be beneficially combined into a hybrid equipment offering cipher integrity and improved audio quality.

Duplex systems working over 2-wire circuits are difficult to implement on links requiring synchronization, particularly if circuit delays are to be kept to a minimum. For these reasons t.e.s. scramblers are likely to be superseded when other equipments offering similar or better security over audio band circuits become available. Digital revertive synchronization systems could offer delay-free duplex communications and can be automatically initiated, thus providing the user with the ease of operation he requires.

It can be seen that the central problem delaying meeting the user requirements is that of low-bit-rate a/d and d/a voice reproduction.[13] Developments of various systems based on delta modulation,[14] 2-bit p.c.m.,[15] and hybrid vocoders are current. One interesting approach being researched at Bath University relies in essence on the signalling of the number of zero crossings that the audio signals make in unit time, and syllabic amplitude information.[16] Transmission of redundancies is avoided, thus permitting bit rates as low as 4·8 k bit/s, but as these redundancies are reintroduced at the receiver voice quality is maintained.

The ciphering of a bit stream presents relatively few problems compared with those of providing good voice quality from digital signals running at 2·4 k bit/s. Future trends must include developments in this area, and should include research into analogue systems with analogue k.g.s that would overcome the security limitations of scrambler systems. In the shorter term crypto-vocoders will be developed that more closely reflect the user requirement. New methods of voice analysis and synthesis are being researched at Cambridge, and if any single factor regarding future developments is significant it is probably in this complex area of low-bit-rate synthesized speech and crypto-vocoders.

8 Conclusions

The use of voice coders has increased markedly in the last decade. Their increased efficiency, smaller size and lower cost has been made possible by the availability of i.c. logic components. Microprocessors and l.s.i.s have enabled coding systems to be implemented that possess a security integrity level that was previously unattainable. Voice coders are now produced by over forty companies in more than eleven countries, and the requirements of Governments and commerce continue to increase.

The security integrity of voice coders for use over normal audio band circuits will improve when low-bit-rate a/d and d/a problems are solved, allowing voice ciphers to be used in place of t.d./f.d. scramblers. In the interim period, possibly four or five years, crypto-vocoders will provide expensive but effective speech secrecy, with analogue output t.d./f.d. scramblers providing adequate cover time for most audio band applications.

It is thought that voice coding circuits will be a standard integral part of communication equipments within a few years, leading to better speech quality and operability than that available from present day appliqué units. Developments in analogue signal analysis and synthesis could provide a breakthrough into analogue speech ciphers in time for them to be incorporated into the crypto-communication equipments foreseen above.

Voice coding plays a growing part in the provision of communication security, now considered to be as important to businesses as it is to government communications. Research into the complexities of the human voice and its successful synthesis, coding and identification is widespread. Future developments in voice ciphers may well be cross-fertilized by work in other areas of analogue signal processing, to the benefit of both. Voice coding is becoming a subject as significant as textual coding has been over the centuries since a Roman Caesar devised a substitution alphabet.

9 References and Bibliography

1 Kirchofer, K. H., 'Secure voice communications cryptophony', *Intl Defence Rev.*, September 1976.
2 Kahn, D., 'The Codebreakers' (Macmillan, London, 1967).
3 French, R. C., 'Speech scrambling and synchronization', *Philips Res. Rep.*, Supplement No. 9, 1973.
4 Gallois, A. P., 'Communication privacy using digital techniques', *Electronics and Power*, **22**, pp. 777–80, November/December 1976.
5 Hinsley, F. H., 'British Intelligence in the Second World War', Vol. 1 (HMSO, 1979).
6 Griffiths, J. W. R. and Tomlinson, M., 'An electronically programmable shift register', *The Radio and Electronic Engineer*, **37**, no. 4, pp. 209–11, 1969.
7 McCalmont, A. M., 'Communications security for voice-techniques, systems, and operations', Telecommunications, April 1973, pp. 35–41.
8 Gold, B. and Rader, C. M., 'The channel vocoder', *IEEE Trans. on Audio and Electro Acoustics*, **AU-15**, no. 4, pp. 148–60, December 1967.
9 Kelly, L. C., 'Speech and vocoders', *The Radio and Electronic Engineer*, **40**, no. 2, pp. 73–83, August 1970.
10 Markel, J. D. and Gray, A. H., 'Linear Prediction of Speech' (Springer, New York, 1976).
11 Pless, V., 'Encryption Schemes for Computer Confidentiality', Massachusetts Institute of Technology, MAC Technical Memorandum 63, May 1975.
12 Jennings, S. and Beker, H., 'Cryptographic requirements for digital speech systems', pp. 5–9, RACALEX Lecture Papers, October 1979.
13 Steele, R., 'Delta Modulation Systems' (Pentech, London, 1975).
14 Lamba, T. S. and Faruqui, M. N., 'Intelligible voice communication through adaptive delta modulation at bit rates lower than 10 kbit/s', *The Radio and Electronic Engineer*, **48**, no. 4, pp. 169–75, April 1978.
15 Sakane, F. T. and Steele, R., 'Two-bit instantaneously adaptive delta modulation for p.c.m. encoding', *The Radio and Electronic Engineer*, **48**, no. 4, pp. 187–97, April 1978.
16 King, R. A. and Gosling, W., 'Time-encoded speech', *Electronics Letters*, **14**, no. 15, pp. 456–7, 20th July 1978.

Manuscript first received by the Institution on 6th August 1979 and in final form on 20th November 1979.
(Paper No. 1924/Comm. 194)

UDC 621.374.4: 621.376.9
Indexing Terms: Operations, Delta modulation, Digital filters

A special-purpose delta multiplier

N. KOUVARAS, B.Sc.*

SUMMARY

A circuit is proposed which transforms a delta sequence (d.s.) of a signal into another delta sequence of a double amplitude signal. With the proposed circuit, it is possible to transform the d.s. of half-sums obtained from the conventional delta-adders into a d.s. of the corresponding full sums. This increases the flexibility of the realization of digital filters with delta input–output sequences.

* *Nuclear Research Centre 'Demokritos', Aghia Paraskevi Attikis, Athens, Greece.*

1 Introduction

A circuit which can transform a delta sequence (d.s.) of a signal into a delta sequence of another signal of double amplitude is described in this paper. This special delta multiplier or delta doubler, as it will be called from now on, can be a very useful element in delta signal processing techniques.

A method for the hardware realization of a delta adder (d.a.) for the addition of d.s. was presented in a recent paper.[1] Yet, the output sequence of a d.a. is not a sequence of the full sum of the signals represented by the two-input d.s. but a sequence of their half sum. This inherent property of the d.a. is very convenient for the synthesis of a delta multiplier, which multiplies a d.s. by a positive constant which must be less than unity.[1] This property, however, imposes some restrictions on the synthesis of digital filters, that is, the value of the coefficients must be less than 0·5 and 0·25 for first and second order recursive digital filters respectively.[1] This reduces the variety of possible filter design with desired characteristics. These restrictions can be removed by using the suggested delta doubler. This is possible because the following combinations can be realized by means of this delta doubler: first, a d.a. and a delta doubler in cascade, which constitute a full delta adder; second, a delta multiplier of a constant by a d.s. in cascade with the delta doubler which constitute a delta multiplier of a d.s. by the double value of the constant. Thus, if delta doublers are connected at the addition and multiplication points, whenever necessary, there will be no restrictions for the synthesis of digital filters.

From now on any d.s. will be represented by a capital letter with an index n and considered as for binary levels -1 and 1 or for logic levels 0 and 1 respectively.

2 The Delta Doubler

The circuit above the dashed line in Fig. 1 is described as doubler while the portion below the dashed line is an additional improvement to be described later.

The sequence X_n is the input d.s. and Y_n the output d.s. which is obtained from the last flip-flop of the up–down counter. The output Z_n of the first flip-flop of the counter takes alternately the values 0 and 1 and is therefore an idling sequence or better a zero signal d.s. Y_n and Z_n are added by means of the delta adder DA2 and X'_n is the sequence of the half-sum. As, however, Z_n is a zero signal d.s., X'_n is a d.s. of the half amplitude of the analogue signal corresponding to Y_n. X'_n is subtracted, due to the inverter, from X_n by means of the delta adder DA1 and E_n is the d.s. of their half-difference. As long as the value of E_n is 1 the counter is on up-counter mode and when the E_n value is 0 it is on down-counter mode. Thus the circuit operates as a comparator for X'_n and X_n while E_n corrects Y_n at each clock pulse so that X'_n tends to approximate X_n or otherwise, X_n is X'_n plus an error

0033-7722/80/040156+02 $1.50/0

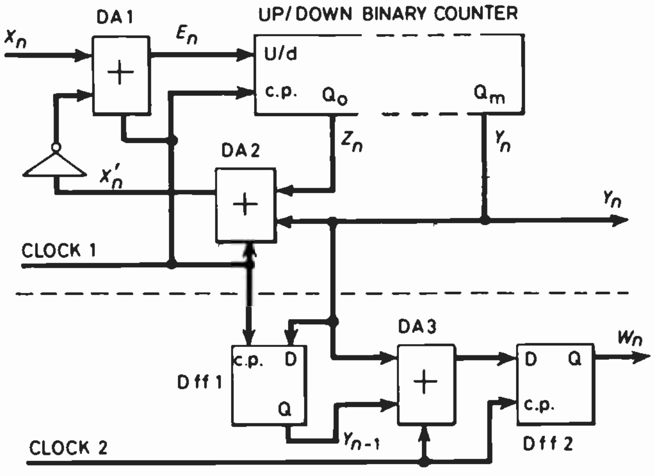


Fig. 1. Delta doubler.

sequence E_n. This, however, means that Y_n is the sequence of the double amplitude signal of that of X_n plus an error.

It should be stressed here that the slope of the analogue signal represented by X_n must not exceed the half of the maximum permissible slope, in order to avoid overloading of the output sequence, Y_n. With this restriction, only two stages are necessary for the up–down counter. The counter operates as a two-stage up–down counter, even if it employs more than two stages, as only the first and the last stages are exploited (Fig. 1). Yet, more stages in the counter prove useful when, during a certain small time interval, the input d.s. happens to lead to an overloaded output d.s. In this case, the two-stage up–down counter will overflow giving a faculty-overloaded output d.s. This can be avoided by using an up–down counter with more than two stages.

The rate of CLOCK 1 is the same as that of the delta modulator which produces X_n.

For the determination of the r.m.s. value of the quantization noise of Y_n, measurements were done as follows. An exponential delta modulator[2] was employed with circuits for the frequency range 50 to 1000 Hz and with a clock rate of 100 kHz. The filter in the decoder was a low-pass one with cut-off frequency of 3 kHz. The measurements used sinusoidal signals within the range of the delta modulator and for various amplitudes. These measurements have shown that the r.m.s. value of the quantization noise of Y_n is twice that of X_n for the low frequencies and three times as much at the higher frequencies of the range. Moreover, if the clock rate of the delta modulator is doubled (200 kHz) the r.m.s. value of the quantization noise of Y_n is never greater than of X_n at clock rate of 100 kHz.

The oscillations of Fig. 2 show (a) an analogue signal decoded from X_n and (b) the corresponding analogue signal from Y_n. The clock rate of the delta modulator is 100 kHz.

A considerable reduction of the quantization noise is achieved with the addition of the circuit shown below the dashed line of Fig. 1. Here, the clock rate of CLOCK 2 is the same as that of the delta modulator while the rate of CLOCK 1 is doubled. Thus, Y_n is the d.s. of a double signal as well as of double sampling frequency. The circuit shown below the dashed line in Fig. 1 is a sampling frequency correction circuit so that W_n becomes a double amplitude signal d.s. with sampling frequency equal to that of X_n. The inputs of the delta adder DA3 are Y_n and the 1-bit delayed Y_{n-1} but the adder is clocked simultaneously with X_n. At the same time, the output of DA3 is shifted to D–ff–2. It can be easily seen that W_n is a sequence of the half sum of Y_{2k} and Y_{2k+1} ($k = \ldots -1, 0, 1, \ldots$) which are subsequences of Y_n and each bit of them is stretched to the time interval between two clock pulses.

The measurements obtained showed an increase in the r.m.s. value of the quantization noise of W_n of only 50% of that of X_n for any sinusoidal signal within the frequency range of the measurements. The oscillogram (c) of Fig. 2 shows the analogue signal decoded from the output sequence W_n of such a doubler.

More details of the suggested circuit can be provided by the author.

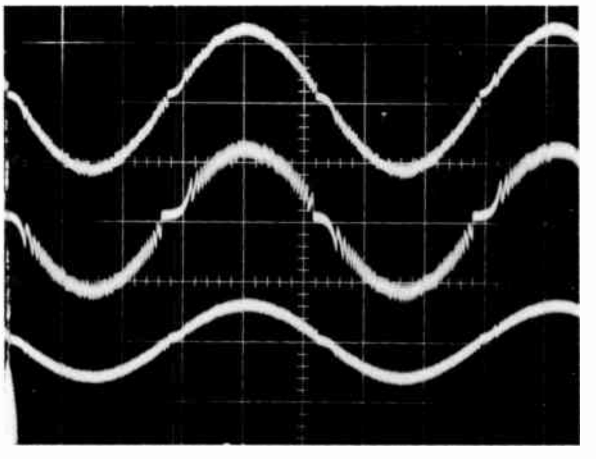


Fig. 2. Analogue signals decoded from delta sequences of delta doublers.
(a) Input signal (X_n).
(b) Output signal from a simple delta doubler (Y_n).
(c) Output signal from an improved delta doubler (W_n).
Vertical scale: 1 V/cm. Horizontal scale: 2 ms/cm.

3 References

1 Kouvaras, N., 'Operations on delta modulated signals and their application in the realization of digital filters', *The Radio and Electronic Engineer*, **48**, no. 9, pp. 431–8, September 1978.
2 Steel, R., 'Delta Modulation Systems' (Pentech, London, 1975).

Manuscript first received by the Institution on 27th June 1979 and in final form on 22nd October 1979.
(Paper No. 1925/CC 320)

UDC 621.3.049.771.14: 621.396.62.53.029.6

Indexing Terms: Integrated circuits, Large-scale integration, Radio receivers

An integrated circuit v.h.f. radio receiver

I. A. W. VANCE, B.Eng., M.Sc.*

Based on a paper presented at the IERE Conference on Land Mobile Radio held at Lancaster from 4th to 7th September 1979

SUMMARY

A radio receiver system for frequency-modulated signals is described which uses a direct conversion (zero intermediate frequency) technique. The characteristics of the system are developed especially in areas of difference from conventional methods.

The realization is shown of the complete receiver in two integrated circuits which form part of a synthesized 30 to 88 MHz equipment. The conclusion is drawn that the combination of direct conversion and large-scale integration is a very powerful means for future development of radio receiver techniques.

* *Standard Telecommunication Laboratories Ltd., London Road, Harlow, Essex CM17 9NA.*

1 Introduction

This paper describes the concept and realization of a large-scale integrated circuit (l.s.i.) radio receiver for frequency modulated signals, capable of operating at up to the v.h.f. bands.

Current receiver developments in both civil and military areas are characterized by an ever increasing complexity in frequency range, modulation characteristics and, especially, operating agility. At the same time there is a need to reduce the size and weight (bearing in mind battery requirements) and to allow handling by operators of limited skill. The solution to this latter problem in other areas has been the l.s.i. chip, but to date the radio equipment field has been resistant to this approach, except for some easily integrated sub-units.

The breakthrough to l.s.i. radio can only be achieved if equipment architectures can be found which eliminate the inherently non-integrable elements of inductors and capacitors, especially the variable variety. Even without the benefits of integration the elimination of these elements would be worthwhile.

The system described below is a receiver of the so-called 'direct conversion' or 'zero intermediate frequency (zero i.f.)' variety. This general form of receiver has the usual superheterodyne advantages of obtaining selectivity and gain at a fixed frequency, but avoids the major problems of the image and associated responses which are present for i.f.s other than zero.

Many forms of direct-conversion receiver have been proposed (e.g. Refs. 1 and 2) over the years but have not found general acceptance due either to their complexity or to the presence of a difficult to control critical element. The integrated circuit approach allows a measure of complexity and it will be shown that the critical elements can be controlled for the system to be considered. The overall concepts here are not new[3] but further development is required to give acceptable receiver performance.[4]

2 The System

A conceptual picture of the operational principle will be given and then some specific points of detail will be examined. Figure 1 shows a block diagram of the receiver system.

2.1 General Description of Operation

The incoming signals are mixed with quadrature local oscillators in two mixers. The oscillator is nominally placed on the same frequency as the received carrier, and two low-pass filters at the mixer outputs then select the difference frequency. If the instantaneous frequency difference is $\delta\omega$ then the two outputs can be considered as sinusoids of the form $a \sin \delta\omega$ and $b \cos \delta\omega$ as shown in the Figure. These are amplified and then differentiated.

0033–7722/80/040158+07 $1.50/0

The process of differentiation produces a further 90° shift and an amplitude linearly dependent on the input frequency. That is symbolically

$$\frac{d(a \sin \delta\omega t)}{dt} = a\, \delta\omega \cos \delta\omega t$$

and

$$\frac{d(b \cos \delta\omega t)}{dt} = -b\, \delta\omega \sin \delta\omega t.$$

These terms are then multiplied in linear four-quadrant multipliers by the undifferentiated opposite channel to form the products

$$ab\, \delta\omega \cos^2 \delta\omega t \qquad (1)$$

and

$$-ab\, \delta\omega \sin^2 \delta\omega t \qquad (2)$$

which when subtracted eliminate the time-variant terms and give

$$ab\, \delta\omega \qquad (3)$$

or, for equal level inputs, $a^2\, \delta\omega$ for $\omega_{in} = \omega_0 + \delta\omega$. That is, a direct voltage proportional to the square of the input amplitude and to the instantaneous frequency difference.

For a negative frequency excursion of the input signal, the sense of the quadrature in the two channels reverses. This is equivalent to a 180° change in one channel and clearly results in the output:

$$-a^2\, \delta\omega \text{ for } \omega_{in} = \omega_0 - \delta\omega.$$

Thus it is seen that the output of the entire system is that of a linear instantaneous frequency discriminator. Since no band limiting or sampling has been performed, and since all the operations are linear, there is no limit to the rate at which the frequency may be varied. Hence, provided the low-pass filters are adjusted accordingly, the system will demodulate f.m. of any index.

2.2 A.G.C. Characteristics of the System

The output has a square-law dependence on the input level which is undesired. It is not possible to employ limiting amplifiers in the two channels as in a conventional i.f. system, as this, by only preserving zero crossings, is effectively a sampling process. For small modulation indices this would imply a signalling speed through the system greater than the sampling rate, in direct contravention of Nyquist's criterion.

The main gain blocks may, however, have automatic level control (a.l.c.) applied to them. Here the output level is detected and a d.c. feedback established to maintain a constant output. The process removes the amplitude-squared dependency by setting:

$$a = k_1 \quad \text{for any } a$$

$$b = k_2 \quad \text{for any } b$$

In general $k_1 = k_2$, where k_1 and k_2 are constants determined by the detector and feedback reference level in the a.l.c. stages.

For a finite response time τ, the a.l.c. feedback network must be allowed some bandwidth (τ^{-1}). Hence input signals to the a.l.c. stages of frequency less than τ^{-1} will propagate round the feedback loop, causing loss of gain or amplitude fluctuations depending on the implementation of the feedback mechanism. However, in a practical receiver system the a.l.c. stages are the main gain blocks and need to have approximately 100 dB of amplification. This cannot easily be obtained with d.c.-coupling, and hence the a.l.c. stages will in any case have a forward path pole causing a low frequency corner with a 6 dB/octave slope to small or unity gain at d.c. Provided the feedback corner is lower than the forward path corner then distortions from this cause can be minimized. The overall effect of these two corners is to generate a frequency response at r.f. which has a zero in the centre. With practical signals of voice or data modulation the effect of this notch is completely negligible, and even with sine wave testing it can be made sufficiently narrow not to be a problem whilst retaining a.g.c. characteristics fast enough to cope with vehicle-to-vehicle communications.

2.3 Sensitivities to Non-perfect Realization

It may be seen by inspection of Fig. 1, and equations (1) to (3), that obtaining an undistorted and spurious-free demodulated output signal is not dependent upon the amplitude balance of the two channels up to the differentiators. That is, no critical balance requirements are placed on the high frequency and variable gain parts of the receiver, due to the cross coupling before the analogue multipliers. If an imbalance exists then exact cancellation of time-dependent products still occurs. The sensitivity with respect to phase imbalance is not so obvious and will be treated below. The frequency-modulated input signal may be represented in the usual way as:

$$a \cos(\omega_0 t + \beta \sin pt)$$

where ω_0 is the carrier frequency, β the modulation index and sin pt the modulation, and where, if $\Delta\omega$ is the peak deviation, $\beta = \Delta\omega/p$. The local oscillator is then:

$$\sin \omega_r t \quad \text{in one mixer}$$

and

$$\cos(\omega_r t + \phi) \quad \text{in the other mixer}$$

where ϕ represents the total phase error in the r.f. channels plus the difference from quadrature in the local oscillators. (It may be noted that the 90° phase shift may be placed arbitrarily in the signal or local oscillator paths, hence it is valid to let ϕ represent the total phase error.)

After mixing and low-pass filtering the in-phase and quadrature signals are:

$$x_i = \frac{a}{2} \sin\{(\omega_r - \omega_0)t - \beta \sin pt\}$$

and

$$x_q = \frac{a}{2} \cos\{(\omega_r - \omega_0)t + \phi - \beta \sin pt\}.$$

Differentiating gives:

$$x_i' = \frac{a}{2}\{(\omega_r - \omega_0) - \Delta\omega \cos pt\} \ \cos\{(\omega_r - \omega_0)t - \beta \sin pt\}$$

(remembering that $\beta = \Delta\omega/p$) and

$$x_q' = \frac{-a}{2}\{(\omega_r - \omega_0) - \Delta\omega \cos pt\} \times \\ \times \sin\{(\omega_r - \omega_0)t + \phi - \beta \sin pt\}$$

The outputs of the multipliers are then:

$$x_i x_q' = \frac{-a^2}{8}\{(\omega_r - \omega_0) - \Delta\omega \cos pt\} \times \\ \times \{\cos\phi - \cos[2(\omega_r - \omega_0)t + \phi - 2\beta \sin pt]\}$$

and

$$x_q x_i' = \frac{a^2}{8}\{(\omega_r - \omega_0) - \Delta\omega \cos pt\} \times \\ \times \{\cos\phi + \cos[2(\omega_r - \omega_0)t + \phi - 2\beta \sin pt]\}.$$

Hence the output after subtraction is:

$$x_q x_i' - x_i x_q' = \frac{a^2}{4} \cos\phi\{(\omega_r - \omega_0) - \Delta\omega \cos pt\}. \quad (4)$$

The first term in the brackets here is the d.c. offset due to being off tune and the second term is the original modulation, directly proportional to the deviation, as expected. These are multiplied by a constant which depends on the phase error. Hence it is seen that phase inaccuracies cause no spurious outputs, but merely contribute a scaling factor which is clowly changing for ϕ near zero degrees. Therefore the r.f. stages, the quadrature phase shift and the a.l.c. gain blocks are insensitive to both phase and amplitude imbalances.

After the differentiators, however, the components must all be balanced, as any imbalances here will cause the $\sin^2$ and the $\cos^2$ terms not to cancel, producing an unwanted spurious term at twice the instantaneous deviation frequency. Similarly, if the analogue multipliers are imperfect and produce outputs other than the wanted product, then these in general will not cancel and will also cause a spurious output. However, these parts of the demodulator are all working at low frequencies (below 10 kHz for a narrow band receiver) with fixed, high level signals from the a.l.c. stages, and in the present realization are to be fabricated on a single chip where balance and tracking with temperature can be controlled. Nevertheless, as with any balancing method, it is to be expected that extreme care will have to be taken if spurious levels are to be held below 40 dB.

2.4 A.M. Rejection

The analysis for an amplitude-modulated signal passing through the system is straightforward but very tedious and will not be presented here. The main results are as follows: when exactly on tune, the symmetrical nature of the a.m. sidebands causes them always to cancel at the output, for any phase of the local oscillator relative to the incoming carrier. When off tune, an output is produced which scales directly with the offset frequency. In this connection the output consists of a d.c. term, representing the mean power of the input, the modulation and components at twice the modulation frequency, twice the offset frequency and the sum and difference frequencies of these. A 20 dB level of a.m. rejection (total power of all output frequencies) can be maintained for an offset frequency up to 0·3 $\Delta\omega$. Automatic frequency control can keep this to a low level.

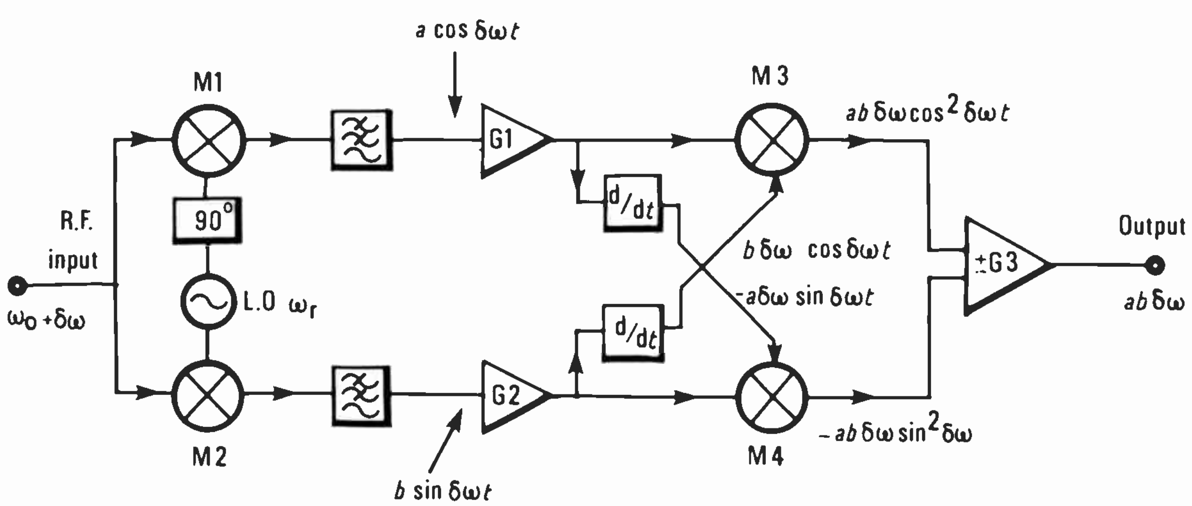


Fig. 1. Principle of basic system.

2.5 Signal-to-noise Considerations

The noise figure of the receiver may be set by a single r.f. amplifier stage before the splitting into two channels occurs, thus defining it as being the same as any other receiver and ensuring correlated noise in the two channels. Moreover, the effective noise bandwidth of the receiver is also the same as a single channel system since each low-pass filter, although half the physical bandwidth of the modulation, receives noise from either side of the local oscillator. Thus the noise density is twice as high in half the bandwidth, giving the same noise power. Of course no noise is received at the image frequency, which is true for the non-zero i.f. case only if adequate filtering is included after the r.f. amplifier. Hence, in the case of a large input signal-to-noise ratio, the usual simplification can be made of regarding the noise as band-limited random-phase variations on the signal carrier. Here, the performance of the present system is clearly identical to that of any other discriminator and the usual f.m. improvement and triangular noise spectrum result.

In the threshold region between the large and small signal cases, the rate of change of output signal/noise is dominated by the a.l.c. stages. In contrast to wideband limiting amplifiers, these show no sharp capture effect, but measure the mean peak level of the signal plus noise. This is the classical d.c. output of an envelope detector, which is treated for example by Panter (Ref. 6, pp. 228–33). The detected d.c. level is within 2 dB of that which would be contributed by the carrier alone down to unity carrier-to-noise power ratio, but below this the noise rapidly dominates. Hence the a.g.c. stages may be considered to be controlled by the signal to the unity power point, in contrast to a limiter, which, since it has no long time-constant averaging associated with it, commences suppressing the signal at higher signal-to-noise ratios (Ref. 6, pp. 455–9). The difference between the two systems depends upon the noise statistics and hence varies from narrowband to wideband modulations. In the communications equipment situation of narrowband modulation the difference is less than 3 dB.

The transfer function of the discriminator following the a.l.c. stages is, as has been seen, square law. This serves to further sharpen the threshold effect, such that overall it is comparable to a conventional limiter discriminator for the narrowband case. (The same mechanism applies to co-channel signals which is discussed in Section 4.)

The analysis for the low signal-to-noise region is not of prime concern for normal voice communication systems, is difficult to perform for a general case, and will therefore not be attempted here.

The system may also be viewed as a power-mean frequency estimator,[5] which allows another method of noise calculation.

3 Practical Realization

The system so far described has been realized, together with some embellishments, on two silicon integrated circuits. The performance of these reflects two developments, firstly the new architecture and secondly the integration of radio frequency components. It is primarily the first item which will be discussed here; the second will be the subject of a later paper.

Figure 2 shows a block diagram of the first chip which contains all the high frequency components. The received signals are amplified and applied to two balanced mixers whilst the on-chip local oscillator is quadrature-phase-shifted with a simple RC network, and

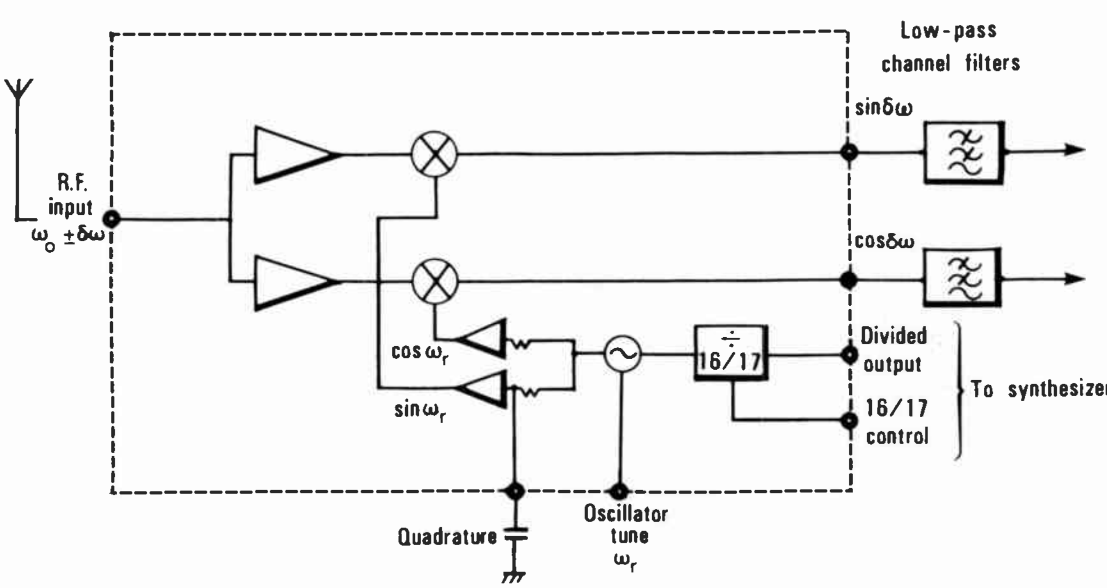


Fig. 2. R.f. integrated circuit.

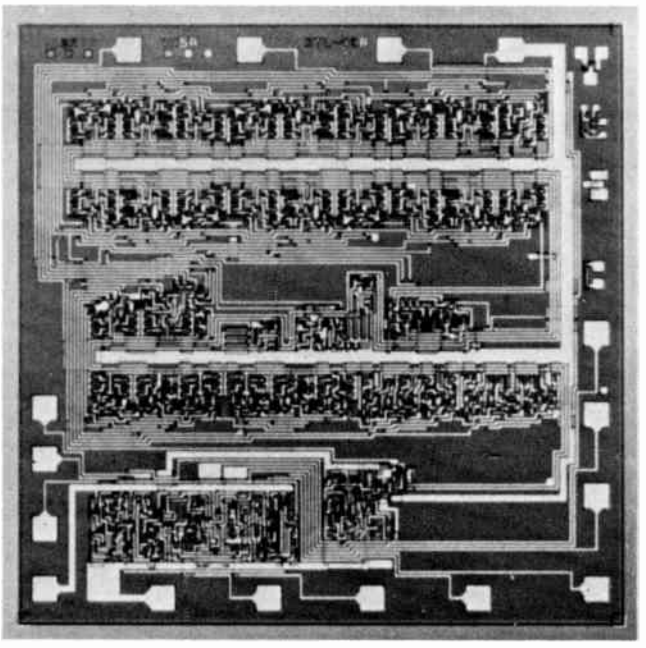


Fig. 3. The h.f. components integrated circuit.

isolated from the mixers with buffer amplifiers. The oscillator is also divided down in a variable ratio divider to interface with a standard synthesizer chip set. Figure 3 is a photograph of the integrated circuit; all the analogue functions are contained in the small block to the lower left, the remainder of the 6·3 mm^2 chip being the divider. This circuit is fabricated in an 800 MHz f_T process and is designed primarily for the 30 to 88 MHz band. The main components could of course be used with a conventional superhet, but three areas are of especial relevance to the present system.

Firstly, since the local oscillator is on-channel, only the reverse isolation of the mixer and r.f. amplifier prevents it being radiated from the antenna. By careful design, including cross-neutralization of the differential r.f. stages, the isolation obtained on-chip is of the order of 80 dB up to 100 MHz. Secondly, the accuracy of the 90° phase shift is maintained with 10° over 30 to 88 MHz, even with the simple system used, again by on-chip compensation of stray elements. Thirdly, in the context that the concept here is l.s.i., it is found that there is no measurable interaction between the analogue and digital portions of the circuit.

The two low-pass filters which join the i.c.s in this realization, and which provide the channel selectivity for the receiver, are of the passive LC type. This approach was chosen over the alternative of active filter types to reduce power consumption and give high dynamic range. The required characteristics for in-band amplitude flatness, group-delay and next-channel rejection are easily met, more easily indeed than with a crystal filter at a high intermediate frequency. A five-section filter achieves 80 dB next-channel rejection.

The low-frequency signal processing is performed in the second i.c., illustrated in Fig. 4. The inputs from the low-pass filters are amplified in the two a.l.c. stages which have 100 dB of gain and a.g.c. range. The differentiators are CR networks, with external capacitors. To improve the balance performance of the analogue multipliers, the non-differentiated inputs are converted to pulse-width-modulated (p.w.m.) signals which then switch a doubly-balanced mixer, whilst the differentiated signals are applied to the mixer's linear input. The conversion to p.w.m. uses an on-chip triangle wave oscillator running at a few hundred kilohertz, and

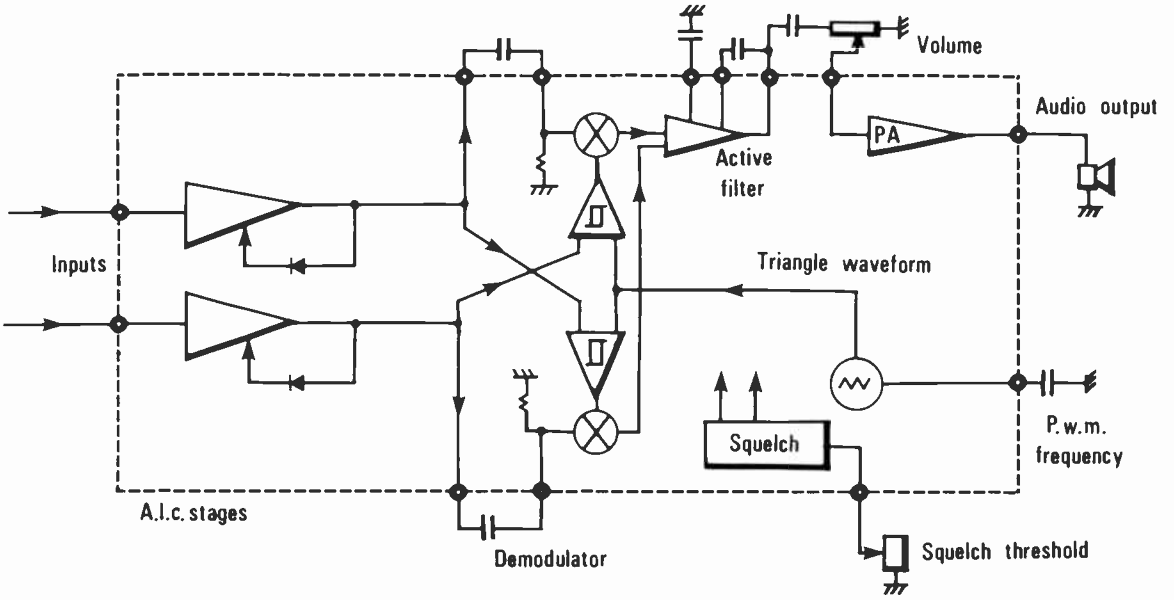


Fig. 4. Signal processing integrated circuit.

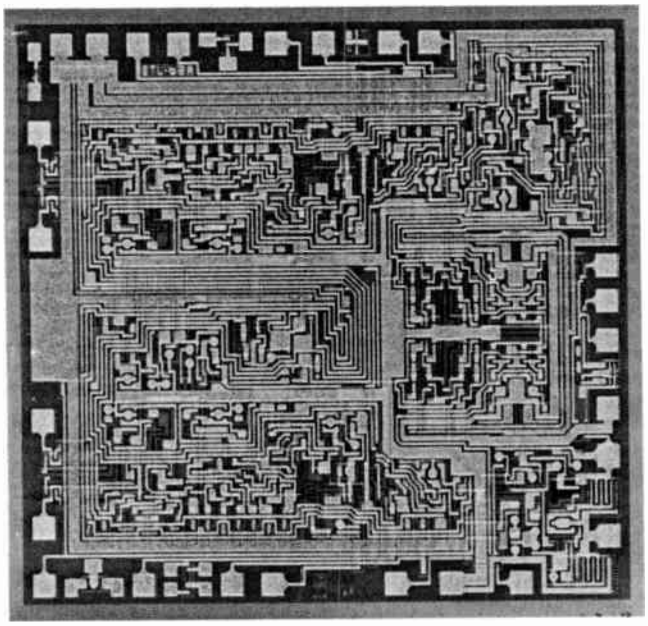


Fig. 5. The signal processing chip.

two comparators, as shown. An active post-detection filter reconstitutes the analogue output, which may be applied to the 20 mW audio output stage which is included (providing sufficient output for a headset). Variable threshold squelch of the carrier level detecting type is also provided. A signal level output is available for 'S' meter or r.f. a.g.c. functions.

Figure 5 shows the signal processing chip itself which is in the standard buried collector bipolar process and is 12 mm^2 in area. Despite some 150 dB of audio gain across the circuit no instability problems are present, although very careful design and modelling was used to ensure this. For example, single-point earthing is used; this can be seen in the metal tracks to the double ground bond pads in the top left-hand corner of the chip.

The worst case unbalance signal at the output from this circuit (which is due to an unmodulated carrier offset from the local oscillator such that the beat note falls in the post-detection filter), is about 36 dB down on the main output. This can of course be controlled by applying a.f.c. It is also a chip-to-chip variable which can be controlled by selection.

The zero-frequency 'hole' in the frequency response is formed firstly by the d.c. blocking capacitor between the two chips and secondly by the d.c. feedback roll-off in the a.l.c. amplifiers.

A receiver incorporating the circuits is shown in Fig. 6.

4 Comparison with Conventional Superhet Receiver

The key improvement obtained over the high intermediate-frequency receiver is the elimination of the first-order spurious response, the image. In a real equipment there will also be a large number of other image related responses due to non-linearities, for example image difference input signals, which are of course also eliminated.

If a sufficiently high dynamic range r.f. section is available then tunable input filtering can be dispensed with altogether or very simple filters can be used. If these are required to track with the local oscillator then they can do so accurately as they are on the same frequency.

The advantages obtained by eliminating the image and associated responses will be greater for equipments requiring large tuning ranges or operating at high input frequencies where the input filtering is a more severe problem.

The channel filters are low-pass and this allows the realization of more complex transfer functions than do the coupling constraints on crystal filters. This is especially important for digital f.m. modes. Furthermore, it is easier to change the filter for different signal formats. As far as the r.f. stages, mixers, and local oscillator are concerned, there are no differences except that the mixer outputs are at a very low frequency which may be advantageous for some configurations.

The overall noise and signal detection properties are very similar to a limiter-discriminator system for the narrowband case, but for wideband modulation the limiter method will show faster capture and a sharper noise threshold. Conversely, at poor input carrier-to-noise levels the new discriminator will deliver a better output signal-to-noise ratio.

4.1 Co-channel Operation

Co-channel performance for the new system is actually measured to be better than a narrowband limiter

Fig. 6. V.h.f. receiver incorporating integrated circuits.

discriminator, if unacceptable distortion of the wanted signal, rather than total output power due to each input, is taken as the criterion. This is due to the smoother changeover which, as expected, is measured to be exactly square-law.

The integration exercise has shown that possible pitfalls such as local oscillator leakage and stability of high-gain single-frequency i.f.s can be readily controlled.

4.2 Power Considerations

Power consumption of the high-frequency components is increased in the new system by the current and local oscillator for the extra mixer. If a high dynamic range receiver is the objective, then the power required for the mixers is proportional to the large signal handling properties needed. Since each mixer only sees half the input level, it will require only half the power, hence the total high frequency circuitry consumption will be approximately the same as a comparable single channel. The power required for the main gain blocks and the signal processing in any system is directly proportional to the frequency response required. In the zero i.f. receiver this is by definition the minimum that can be achieved. The requirement for linearity in the gain stages as opposed to limiting amplifiers does lead to an increased power requirement of about five to one. This implies that limiters have to be used at below 100 kHz to complete: the double or triple conversion that this in turn implies will again consume more power; so it is seen that the zero i.f. receiver can be an optimally low power equipment.

4.3 Integration

The major advantage of the new architecture is that it allows the straightforward realization of the receiver in integrated form, together with the elimination of critical tuning adjustments. Since most of the operation is at low frequencies, printed board layout and screening also become much easier. In general the result is a reduction in the skill necessary to design and assemble such equipments and a consequent increase in reliability and serviceability. From a cost standpoint the investment in the i.c. design is outweighed even for modest production volumes by the savings in assembly, testing, and final size and weight. In this sense integration brings greater benefits to r.f. equipments than to digital or low-frequency analogue where the savings to be made are less.

Electromagnetic compatibility in both directions is improved by miniaturization, as the coupling elements, capacitive and inductive, to the outside world are much smaller.

4.4 Transmitters

Whilst it is not the intention here to cover transmitters, mention must be made of the fact that with the local oscillator on-channel the need for 'i.f. shift' between transmit and receive is removed. This results in an even greater simplification of the total equipment. Modulation can be accomplished directly, via phase modulation or with frequency feedback techniques.

5 Conclusions

A direct conversion radio receiver for frequency-modulated signals has been described. It has been shown that whilst all of its characteristics are not identical to existing methods of realization, no significant penalties are paid by its use and indeed it allows a measure of integration previously unachievable in radio equipments.

It is to be expected that further developments will be made to improve on the system described and that it will be extended to other signal formats. This is to be contrasted with the conventional receiver which has been under continuous development for many years, and may be considered to be fully evolved. The true l.s.i. one-chip receiver may be confidently predicted for the near future, and this in turn opens up new possibilities for systems where equipment cost or bulk have previously been a limitation.

6 Acknowledgments

I am greatly indebted to numerous of my colleagues who have provided inputs on this programme over a period of time, especially M. W. Neale and R. J. Miller.

7 References

1 Weaver, D. K., 'A third method of generation and detection of single-sideband signals', *Proc. IRE*, **44**, no. 12, pp. 1703–5, December 1956.
2 Costas, J. P., 'Synchronous communications', *Proc. IRE*, **44**, no. 12, pp. 1713–18, December 1956.
3 British Patent Specification 1,363,396, 'Sine-cosine Frequency Tracker'.
4 British Patent Specification 1,530,602, 'High Dynamic Range Receiver for Frequency Modulated Signals'.
5 Denenberg, J. N., 'The power mean frequency estimator: another approach to the f.m. detector', *IEEE Trans. on Broadcast and Television Receivers*, **BTR-20**, no. 3, pp. 201–4, August 1974.
6 Panter, P. F., 'Modulation, Noise and Spectral Analysis' (McGraw-Hill, New York, 1965).

Manuscript received by the Institution on 17th April 1979
(Paper No. 1926/Comm 195)

UDC 621.38: 621.3.037.375

Indexing Terms: Display systems, Alpha-numerics, Liquid crystals, Matrix addressing, Character generation

An analysis of possible drive schemes for complex liquid-crystal displays

R. BENJAMIN,
C.B., D.Sc., C.Eng., F.I.E.E., F.I.E.R.E.*

SUMMARY

The paper is concerned with the problem of generating complex patterns on two-dimensional liquid-crystal arrays, without either an excessive number of control terminals or prohibitively complex control circuits. Following a brief introduction to the operating principles of liquid-crystal displays, the initial Section indicates how the problem could be solved using a two-dimensional shift-register/liquid-crystal sandwich—if the technology were available. Succeeding Sections then derive and analyse efficient ways of arranging and driving sets of parallel X and Y conductors to generate desired display patterns. Particular attention is given to the examples of generating arrays of 7×5 dot matrix symbols, and of generating single two- or three-valued waveform displays. Drive sequences within the response speed of the liquid crystals are shown to be feasible for character displays, and to have the advantage of simplicity and tolerance to variations in operating conditions. Mean-squared, slow-response drives are shown to be more widely applicable and to provide superior visual performance in certain conditions, providing that the ratio of the mean-squared white-to-black voltages is adequate.

* *Government Communications Headquarters, Oakley, Cheltenham, Gloucestershire GL52 5AJ.*

1 Introduction to Liquid-Crystal Displays

A representative type of liquid-crystal material takes the form of long 'rod'-like electrically-polarizable molecules. A thin film of this material is normally enclosed between two closely-spaced parallel transparent sheets. Surface affinity effects will then cause the rods to favour alignment parallel to the two surfaces. This surface alignment can be made directionally determinate and almost complete, if it is given a directional guide by treating the surfaces with a suitable deposit, so as to produce a set of fine parallel ridges along the chosen direction. Now, in one form of array, the ridges on the 'bottom' surface of the film are made orthogonal to those at the 'top'. The rods in consecutive strata of the film will then progressively change their alignment, so as to produce a smooth rotation of direction through 90° between the top and bottom surfaces. Thus the film will also act as a polarization-converting light guide; each surface, if illuminated with light of the appropriate polarization, will then deliver it at the opposite surface transformed to the orthogonal polarization. This array may therefore be sandwiched between two sheets of conventional polarizing material, each aligned to match the orientation of the adjacent rods. The configuration will thus be 50% transparent to non-polarized light (by accepting the 50% of the intensity aligned with the polarization of the top surface).

Furthermore, placing a mirror below the bottom polarizer will return the same 50% of the light back to the top (and to its original polarization) thus yielding (ideally) a 50% reflectance. If however an electric field of sufficient strength (of either polarity) is applied across the array, it will override the surface-affinity effects (except for a very thin boundary layer) and align the rods parallel to itself (the electric field), i.e. normal to the containment surfaces. It will thus destroy the polarization-rotating character of the liquid-crystal. Hence we are then left with the two orthogonal external polarization-selective filters, thus giving (ideally) zero reflectance. (See Fig. 1.)

Alternatively, the external polarization filters may be parallel to each other, so that there is 'full' 50% reflectance in the presence of a field and zero reflectance, in its absence. For convenience, this scheme, with full voltage for 'white' and zero voltage for 'black', will be assumed hereafter, and the (50%) limiting reflectance achievable will be referred to as 'full white', 70% thereof (35% reflectance) as 0·7 of full white, etc. In practice, polarization rotation is almost fully maintained at applied fields up to, say, roughly ±3 V, but it is almost completely destroyed by fields exceeding, say, ±6 V. This switching between 'black' and 'white' is the mechanism whose control we shall be discussing in this paper.

For standard 'twisted nematic' materials of the type described, the switching time is some tens of milliseconds

0033–7722/80/040165+12 $1.50/0

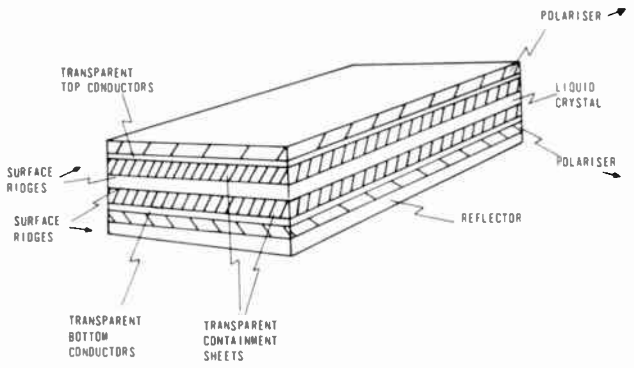


Fig. 1. The constituents of a controllable l.c.d. array.

and, for driving waveforms of a shorter Nyquist interval, the switching drive may be taken to be the mean-squared voltage. However, some equivalent materials are now available where the switching time can be reduced to less than one millisecond. The paper will be concerned, in part, with exploring the possibility of profiting from such faster response times.

Since liquid-crystal displays (l.c.d.s) operate by electrically modifying the reflectance of each spot-element, they can maintain the contrast ratios achieved over almost indefinite variations of ambient light. However, this same characteristic also entails two disadvantages. One—which does not directly concern us here—is that the display depends on the presence of an external source of light. The second disadvantage is that, even if 100% reflectance were achievable (rather than, at best, 50%), the peak brilliance of the display could not approach that of the ambient illumination, unless that reflectance were sustained for a substantial proportion of each viewing 'frame' time. This frame time should be rather less than the viewer's persistence of vision—and in practice lies between 15 and 60 ms. This requirement for sustained activation rules out the extensive time-sharing by which, for instance, each spot on a television display is driven for only about $2/10^6$ of the total time. (Some limited time sharing, exploiting the material's relaxation time, may however be feasible.)

2 Driving Schemes for L.C.D.s

2.1 Individual Cell Drives

For a single-spot l.c.d. indicator light, this requirement for sustained activation is clearly no problem. For a small—e.g. linear—array of spots, a set of parallel drives can often be provided without much trouble. With a slightly larger array, the number of parallel external inputs may be troublesome. However, this difficulty may then be overcome by means of a shift register, converting a single rapid serial input into n sustained parallel outputs. (See Fig. 2(a).) If half-tones intermediate between 'black' and 'white' have to be covered, an analogue c.c.d. shift register could be used for this purpose. With present technology, each output of such a shift register would probably need an f.e.t. emitter-follower or other impedance-converting buffer amplifier. However, further development might permit the c.c.d. array and l.c.d. array to be combined back-to-back in a single multi-layer sandwich, so that the voltage generated by the charge on each c.c.d. shift-register stage also directly acts as the drive to the back-surface of the l.c.d. array. This might be achievable either directly or—with yet more layers in the sandwich and more processing stages in manufacture—via an intermediate stratum of emitter followers. This scheme is illustrated, in grossly over-simplified form, in Fig. 2(b). Even the order of the layers might have to be changed, or multiple functions might have to be combined in one layer, in order to meet competing optical, electrical and physical requirements.

Yet larger, two-dimensional arrays could be made up of a set of, say, m parallel line arrays, each driven by the associated shift register as above. The initial drives to these m shift registers could be provided as a set of m parallel inputs. Alternatively, these too could be serial/parallel converted, by means of an m-stage shift register, which accepts groups of m serial inputs, and then transfers these in parallel to the input terminals of the m line registers (of n stages each). (See Fig. 2(c).) This, again, could be implemented (in the same number of strata) in a composite multi-layer sandwich.

If and when this or an equivalent technique has been shown to be

(a) feasible,
(b) reliable,
(c) economical,

it is likely to become the normal way of generating liquid-crystal displays, at least for complex two-dimensional patterns of arbitrary structure. In the

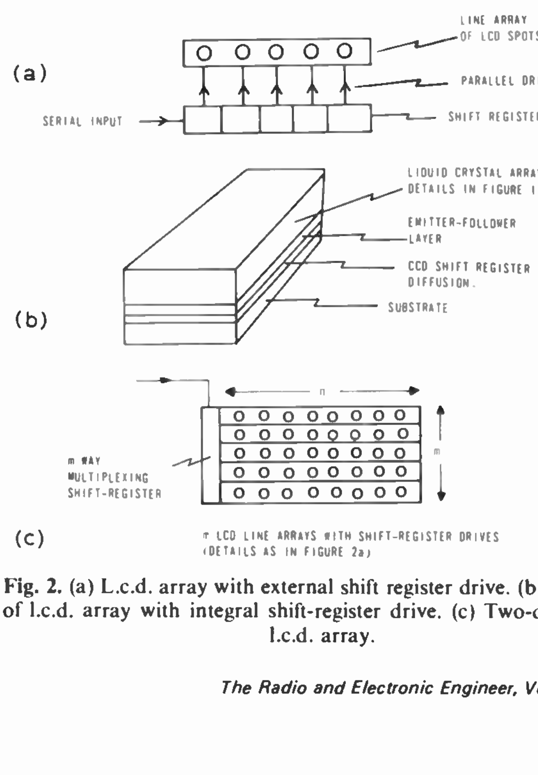


Fig. 2. (a) L.c.d. array with external shift register drive. (b) Schematic of l.c.d. array with integral shift-register drive. (c) Two-dimensional l.c.d. array.

meantime, however, there are requirements for two-dimensional l.c.d. displays of more defined structure, which should be obtainable by the intelligent use of existing technology, with crossed X (row) and Y (column) conductors on, say, the front and back planes, respectively, of the l.c.d. array. The l.c.d. element at each mode will then experience a drive defined by the (sign-independent) voltage difference between the corresponding X and Y conductors. This is the configuration—and problem—discussed in the remaining Sections of this paper. (See Ref. 1.)

2.2 'Row-and-Column' Drives

We can say at the outset that, with this configuration, we should not attempt to create any half-tones (at least not deliberately!). Furthermore, we can state that (say) the vertical stack of V row-drive conductors (for V different Y values) will have to be driven by V distinct (and, in general, mutually 'orthogonal') waveforms. These waveforms will then produce a succession of distinct patterns across the stack of V row-drive wires (one per 'Nyquist interval' of the drive waveforms). For each such instantaneous pattern of row drives, each vertical column conductor will have to be driven at an appropriate voltage, determined jointly by this row-drive pattern and by the desired black/white pattern along the given vertical line. (See Fig. 3.)

In principle we should expect the optimum pattern of V vertically-stacked row-drive sequences to be a function of the V desired horizontal black/white sequences (and of the instantaneous patterns on the H horizontally stacked column-drives), concurrently with the converse relation in the other coordinate. However, it appears that such a scheme would be immensely complex when dealing with arbitrary and frequently-changing black/white 'pictures', and that it would give only a marginal benefit. Hence a simpler scheme is postulated, where one set of drive waveforms is pre-determined (for the moment assumed to be the V vertically-stacked row drives). The other one (the H horizontally-stacked column-drive waveforms) will then be 'designed' to generate the specific black/white pattern required.

It is clearly impossible for an instantaneous set of V row drives and H column drives to achieve a perfect match to an arbitrary pattern of HV black or white spots. However, having now accepted that the row drives shall be pre-determined, it would appear logical to regard the desired pattern of blacks and whites along any column as a 'feature', and to choose the sequence of instantaneous column-drive voltages so that its interaction with the sequence of row-drives will optimize the display of that feature. A suggested criterion for selecting the *instantaneous* column drive would then be the maximization of the mean-squared white/black ratio, *averaged over the column*. This will normally also give a good approximation to maximizing the mean-squared white/black contrast ratio, averaged both over the total pattern of H columns and over the 'frame' repetition time of the row drives (matched to the human persistence of vision).

3 The Choice of Drive Sequences

3.1 Selecting the Pre-determined Row-drive Sequence

We now must consider the choice of the V row-drive waveforms. If they are to permit all possible two-dimensional picture patterns, these waveforms must be mutually orthogonal. If they are to operate equally well with all possible vertical (column) patterns of 'blacks' and 'whites', the obvious choice is to provide the whole range of such patterns in sequence. However, there would then be 2^V such binary patterns. Unless V is very small, this sequence—if realizable at all—would last longer than the persistence of vision, and so would lose much of its purpose; indeed the sequence would probably last longer than the time for which an unchanged object picture is required to be generated! Hence, in many applications, we would want to select a suitable sub-set of the total range of possible vertical patterns. One sub-set which might come to mind is clearly that actually arising on the H columns of the current picture. If this is not sufficiently large, random or orthogonal, it could be 'diluted' with other patterns, either selected arbitrarily or selected, say, because of their occurrence in wanted patterns in the (recent?) past. Indeed, many applications (such as simple alpha-numeric displays) would be inherently restricted to a modest range of pre-determined vertical-line black/white patterns.

However, further consideration suggested that, if this approach were indeed the right one in general, it must also be the right one in the particular case of the oscilloscope application, where each vertical line will normally have only a single white spot on it. However, it turned out to be clearly quite inappropriate in this case. This led to a general reconsideration, and to the initial conclusion that the drives should comprise, sequentially, the set of vertical voltage patterns, across the stack of V horizontal conductors, which comprises all the different combinations of $V/2$ '+' drives and $V/2$ '−' drives. However, for even numbers of whites this then produces some cases where the whites and blacks are both shared

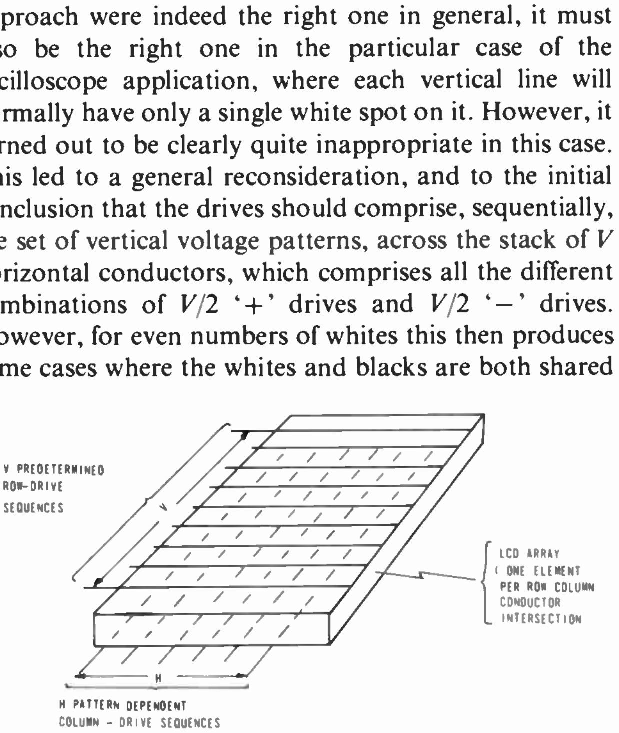


Fig. 3. L.c.d. array with row-and-column drive.

equally between the two drives. To avoid this—clearly unsatisfactory—condition, it was finally concluded that we should use only odd values of V, each split into $(V+1)/2$ '+' drives and $(V-1)/2$ '−' drives—and vice versa. This set comprises 'only' $2V!\bigg/\left(\frac{V+1}{2}\right)!\left(\frac{V-1}{2}\right)!$ members. Even this rapidly becomes excessively large, with increasing V. However, once again, a suitably randomized sub-set of this sequence could then be used with little loss of effectiveness.

3.2 The Corresponding Column-drive Requirements

With this scheme, one of the halves (the + drives or the − drives) will always contain the majority of the desired 'whites' in a given column and hence also the minority of the desired 'blacks'. Conversely, the row drives of the opposite sign must then contain the balance, comprising a minority of the 'whites' and the majority of the 'blacks' in the column. It is then easily seen that, for maximum instantaneous mean (or mean-squared) white/black contrast, the instantaneous column drive should be the opposite to the row-drive containing the majority of the 'whites'. Thus the majority of 'whites' (and a minority of 'blacks') will experience the full white-drive voltage difference, and the majority of the 'blacks' (together with a minority of the 'whites') will experience zero drive voltage, corresponding to 'full black'. (It is in fact possible that either the 'blacks' or the 'whites' (but, with V odd, never both) may be shared equally between the instantaneously + and − rows. In that case, the unequally split colour will of course determine the column drive.)

Given a complete or representative set of these quasi-equally shared row-drives in the frame sequence, all whites will experience the same ratio of 'full' to 'zero' excitations (with 'full' excitation predominating), and all blacks will experience a common ratio of 'full' to 'zero' excitations (with 'zero' predominating). Hence, when integrated over the frame sequence, each column feature will finish up with visually uniform white and black levels. However the actual levels so achieved will depend on the total number of whites and on V, the total height of the column.

3.3 Advantages of the Scheme

It was found that this scheme, using only odd V values and 'quasi-equal' numbers of + and − in its drive patterns, compares favourably with a row pattern drive sequence comprising all the possible column patterns as follows:

(i) It permits a shorter drive sequence.
(ii) Hence, in appropriate applications, it permits 'instantaneous response' of the crystal material to the drive.
(iii) It achieves a higher white/black contrast.
(iv) It minimizes the spread of 'black-level' and 'white-level' reflectances between different black/white column features.

3.4 Relation of the Drive Rate to the Crystal's Relaxation Time

All that has been said so far—and much of the analysis below—is independent of the relation between the Nyquist interval of the driving waveforms and the relaxation time of the crystal material. However some aspects of the performance attainable are dependent upon it. Several authors (see Refs. 2, 3 and 4) have considered systems in which the voltage across the l.c.d. cells has a Nyquist interval short compared with the relaxation time of the crystal material. If the crystal is polarizable but not intrinsically polarized, the resultant alignment of the crystal material to the field is then a simple function of the mean-squared voltage across the field, and the optical reflectivity is derived from this resultant 'averaged' crystal alignment.

In this paper the author has considered the mean-squared 'slow-response' drive discussed above. First of all, however, he has set out to find ways of maintaining each 'instantaneous' drive pattern for a period exceeding the relaxation time of the crystal, and relying on the persistence of vision in the observer's eye to perform the required averaging. Where this is possible, it should permit quite considerable economies in the cost, complexity and power consumption of drive generation and distribution, and in some circumstances it can also result in a display of improved visual performance. The analysis below treats the latter direct-response case, although it is later adapted, to the optimization of the mean-squared mode of operation.

4 Performance Analysis

4.1 General Analysis

To analyse the performance achievable, let the total number of white elements in a column feature be W, and let the total number of black elements be B (i.e. $W+B=V$). Furthermore, let any particular row-drive pattern (in association with the appropriate column drive) engender full reflectance for a majority w of the white elements and a minority b of the black elements in the given column. Hence

$$w > W/2 \text{ and } b \leqslant B/2 \quad (W \text{ odd hence } B \text{ even})$$

or

$$w \geqslant W/2 \text{ and } b < B/2 \quad (W \text{ even hence } B \text{ odd}).$$

Within a full set of row-drive patterns, there will always be just one drive fitting any one specific pattern of w whites and b blacks (together with its inverse, with + and − interchanged), for a given black/white column feature. The total number of such patterns arising is however given by the number of ways of selecting w whites and b blacks out of W whites and B blacks, i.e.

$$N(w) = \frac{W!}{(W-w)!\,w!}\;\frac{B!}{(B-b)!\,b!}.$$

Note that two variants of the right-hand term have normally to be used, entailing $b_1 = (V-1)/2-w$ and $b_2 = (V+1)/2-w$, since each drive splits V into two (slightly) unequal halves. $N(w)$ is then the sum of these two contributions.

Since the patterns of type w will be spread equally over the W wanted white elements, they will contribute full-white reflectance to these elements for w/W of the time. At the same time, they will contribute an average reflectance of b/B to each of the elements intended to be black. (Fortunately we have ensured, by our choice of w, that $w/W > \frac{1}{2} > b/B$—although one at a time can become equal to $\frac{1}{2}$.)

The total white and black reflectances will then be the resultants arising from the summations of the above functions for all relevant values of w, weighted by their relative frequencies of occurrence. Calling these L (light) for the whites and D (dark) for the blacks, they are:

$$L = \sum_w \frac{w}{W} N(w) \Big/ \sum_w N(w); \quad D = \sum_w \frac{b}{B} N(w) \Big/ \sum_w N(w).$$

This is essentially a short-hand notation, since in fact each summation is the sum of two constituents, one being a function of (w, b_1) and one of (w, b_2). The limits of the summations (for both L and D) are as shown in Table 1.

Table 1

Limits of w for the summations for L and D

		Upper limit	Lower limit
W even	b_1	Smaller of W or $\frac{V-1}{2}$	$\frac{W}{2}$
	b_2	Smaller of W or $\frac{V+1}{2}$	$\frac{W}{2}+1$
W odd	b_1	Smaller of W or $\frac{V-1}{2}$	$\frac{W+1}{2}$
	b_2	Smaller of W or $\frac{V+1}{2}$	$\frac{W+1}{2}$

4.2 Properties of the L, D (light, dark) Distribution

If a white-on-black pattern involves W whites in a column of size V, then a black-on-white pattern involves $(V-W)$ whites. However, it can be shown that:

if W, V results in

$$L = 0{\cdot}5 + l, \qquad D = 0{\cdot}5 - d,$$

then $(V-W)$, V results in

$$L = 0{\cdot}5 + d, \qquad D = 0{\cdot}5 - l.$$

This implies that black-on-white always gives a better contrast if $l \geqslant d$. This condition is in fact always satisfied if $W \leqslant V/2$. Hence, if the pattern occupies less space than the background, it should normally be best to generate it in the black-on-white form.

The above skew-symmetry condition between W and $(V-W)$ can also be used to economize in computation, as can the further relation that any even value of W gives the same L and D values as $W+1$, the next odd value. These relations, as well as practical values of L and D achieved, are illustrated in Table 2.

Table 2

'Light' and 'dark' levels (L and D) for various total numbers of whites (W) and vertical column weights (V)

$W =$		0, 1	2, 3	4, 5	6, 7	8, 9	10, 11
$V = 1$	$L =$	1					
	$D =$	0					
3	$L =$	1	0·67				
	$D =$	0·33	0				
5	$L =$	1	0·70	0·60			
	$D =$	0·40	0·3	0			
7	$L =$	1	0·71	0·63	0·57		
	$D =$	0·43	0·37	0·29	0		
9	$L =$	1	0·72	0·64	0·60	0·56	
	$D =$	0·44	0·40	0·36	0·28	0	
11	$L =$	1	0·73	0·65	0·61	0·58	0·55
	$D =$	0·45	0·42	0·39	0·35	0·27	0
∞	$L =$	1	0·75	0·69	0·66	0·64	0·62
(a)	$D =$	0·5	0·5	0·5	0·5	0·5	0·5
∞	$L =$	0·5	0·5	0·5	0·5	0·5	0·5
(b)	$D =$	0	0·25	0·31	0·34	0·36	0·38

Notes: (a) represents white-on-black: (b) represents black-on-white (i.e. 'W' should be interpreted as 'B'). Both (a) and (b) assume that the number of pattern elements is very small compared with V.

When $V \gg W$ $(V \to \infty)$, the function of B can be neglected in computing $N(w)$, and D reduces to 0·5, but

$$L = \sum_{x=0}^{W} \frac{w}{W} N(w) \Big/ \sum_{x=0}^{W} N(w), \text{ broadly as before.}$$

However,

$$\sum_{x=0}^{W} N(w) \text{ simplifies to } 2^W.$$

Here the parameter x defines the range of summation, and w is a simple function of x, as follows:

$$w = x \quad \text{when } x \geqslant W/2$$
$$w = W - x \text{ when } x < W/2.$$

This result can either be derived directly from the various $+/-$ ratios within the W wanted 'whites', or it can be deduced from the previous equations and Table 1 simplified by the fact that the limits of summation are now determined solely by W. The reflectances for $V \gg B$, when aiming at a black-on-white pattern, are complementary to the above, giving a white level of 0·5 and a black level which is 'full white' minus the white level for white-on-black. (See the bottom two rows of Table 2.)

Quite generally, the Table shows that the contrast, L/D, is best if the wanted pattern is black-on-white (i.e. $B < W$), together with a small V and small B/V. If we desire a small spread of the L and D values, we must add the further conditions that $B > 1$, $W > 1$.

4.3 The Effect of Selective Weighting

We may note that, in each V, W combination, the principal degradation of contrast arises from the w, b combination with the largest b/w. If we therefore applied zero column drive whenever this worst-case row-drive pattern arises, we would inhibit any reflectance from 'black' or 'white' elements at that time, and so would improve the contrast quite markedly. This would however be achieved only at the expense of a substantially reduced mean 'white' level, and a substantial spread of this white level between different W. However, any column drive intermediate between full-on and zero will maintain the same contrast for this particular combination of w and b, but will reduce its contribution to the integrated light and dark levels L and D. Hence as a compromise, we might, say, select the drive levels, for these combinations of wanted column feature and vertical drive pattern, which make the white levels identical for all column features. This is illustrated for $V = 3$, 5 and 7 in Table 3. In this Table, the entries under w, b define the condition where drive amplitudes other than $\pm$ unity are required, the entries under wt give the 'weight' (i.e. relative reflectance achieved with the drive amplitude used on these occasions), and L, D give the resultant 'light' and 'dark' levels relative to 'full' (i.e. $\sim 50\%$) reflectance.

For the case of $V = 5$ further figures are listed in Table 4 for reducing the weights from a total of 5 to 4, of relative reflectance 1, 0·7, 0·33, 0, to permit them to be identified by two binary digits. Furthermore it lists an alternative (with weights of 1, 0·67, 0·25, 0) yielding a lower uniform 'white' reflectance, coupled with a better contrast ratio, L/D. It may also be noted in passing that, even with selective weighting, it still remains undesirable to use even-numbered values of V. (For $V = \infty$, weighting will be considered in Section 5.4 and Tables 8 and 9.)

Table 3

Weightings (wt) of column drives for uniform light (white) levels L

V	W	w, b	wt	L	D
3	0	all	0	—	0
	1	1, 1	1/2	0·67	0·17
	2	—	—	0·67	0
	3	—	—	0·67	—
5	0	all	0	—	0
	1	1, 2	1/3	0·6	0·2
	2	1, 1	2/3	0·6	0·23
	3	2, 1	3/4	0·6	0·22
	4	—	—	0·6	0
	5	—	—	0·6	—
7	0	all	0	—	0
	1	1, 3	1/4	0·57	0·21
	2	1, 2	1/2	0·57	0·26
	3	2, 2	7/12	0·57	0·26
	4	2, 1	7/9	0·57	0·25
	5	3, 1	5/6	0·57	0·24
	6	—	—	0·57	0
	7	—	—	0·57	—

5 Application to Practical Requirements

5.1 Application to 7 × 5 Dot-Matrix Character Displays

The foregoing examples show that acceptable drive conditions are available, for controlling the column features of a 7×5 dot matrix. However, much better uniformity, contrast and performance margins could be obtained if the predetermined and pattern-dependent drives were interchanged, so that the matrix would be defined in terms of 7 *row* features of 5 elements each.

If we now consider a representative implementation of the full set of alphabetic and numerical characters and punctuation marks together with the more common mathematical and 'house-keeping' symbols (see, for instance, Fig. 4), we find that this set of symbols uses only about 20 of the 32 possible row patterns of blacks and whites. Furthermore 6 of these are the inverses of 6 others (blacks and whites interchanged) and each such pair can be covered by a single-row drive sequence (with + and − interchanged). Hence only 20 (reducible to 14) row drive sequences need be used in pattern generation.

Furthermore the number of ways of splitting the 5 elements of a row into 3 '+' and 2 '−' is only 10. Thus, by inverting an appropriate set of 5 of these 10 into 3 '−' and 2 '+', all requirements can be met by a sequence of only 10 column-drive patterns, each of 5 '+' and 5 '−' drives (see, for instance, Table 5). Thus, for each of the 20 (or 14) row features, we need only memorize a sequence of 10 '+' and '−'. However, if we 'weight' the drive amplitudes (see Sect. 4.3), we must also memorize the appropriate weighting factors, thus requiring two more bits per element in the sequence. Each character will thus define the appropriate permutation of 7 (of 20 possible)

Table 4

Alternatives to Table 3, for $V = 5$ only, (a) for fewer weights, (b) additionally for better contrast but poorer 'white' level

W	w, b	wt	L	D	wt	L	D
0	all	0	—	0	0	—	0
1	1, 2	1/3	0·6	0·2	0	0·4	0·1
2	1, 1	0·7	0·61	0·23	0	0·4	0·1
3	2, 1	0·7	0·58	0·22	0·25	0·4	0·07
4	all	1	0·6	0	0·67	0·4	0
5	all	1	0·6	—	0·67	0·4	0

Table 5

Sequence of ten column drives covering all 2/3 combinations of $\pm$ and $\mp$, with zero mean

+	+	−	−	−
+	−	+	−	+
−	−	−	+	+
−	+	+	+	−
+	−	+	−	−
−	+	+	−	+
−	−	+	+	−
+	−	−	+	+
−	+	−	−	+
+	+	−	+	−

row features, and each of these in turn will identify the corresponding 10-long row drive sequence (specified by 10 × 3 bits). The practical implementation of this scheme is shown in broad schematic form in Fig. 5, and in slightly more detailed form in Fig. 6. Note the very modest storage requirements involved. The small number of patterns in the drive cycle also avoids any visual problems due to flicker, or any engineering problems due to the response time of the liquid-crystal elements. Furthermore, this sort of drive cycle clearly allows ample time for any multiplexing in the generation and distribution of the various row and column waveforms.

5.2 Mean-squared 'Slow-response' Drives

Since the L and D values on Table 2 are means of a sequence of 1's and 0's, they are equally also mean-squares. It will be noted that, for $V = 5$, the individual columns all provide L/D ratios of 2·33 or better. Such a ratio might perhaps suffice for switching between 'full black' and 'full white', when integrated over the frame sequence. In fact, twisted nematic crystal materials with relaxation times of as much as 100 ms are readily

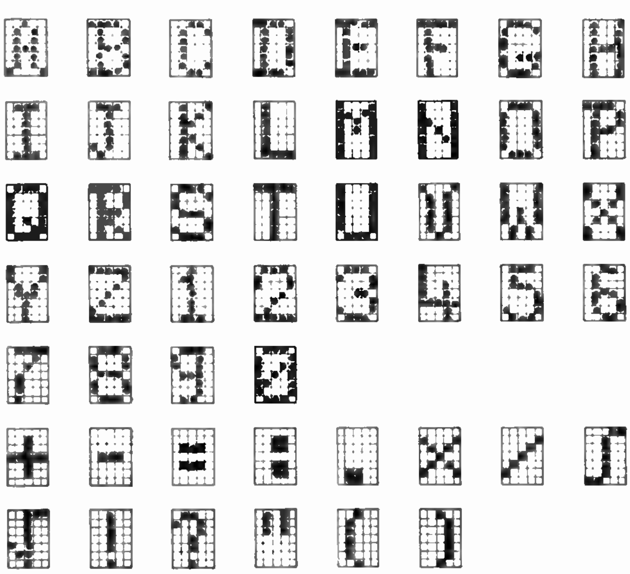


Fig. 4. Examples of 7 × 5 dot-matrix symbols.

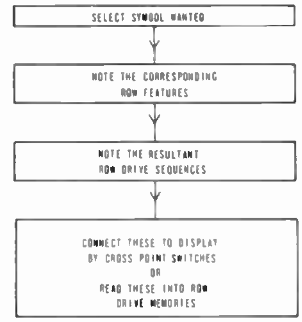


Fig. 5. Simple schematic of row-drive pattern generation and control.

available. Hence this integration will be achieved, if the cycle time is made no more than, say, 40 ms, corresponding to a clock period of 4 ms. Indeed, even when matching to much faster crystal response speeds, waveform generation and distribution would still, clearly, meet no circuit response-time/bandwidth problems.

The foregoing would be of only academic interest if we were unavoidably constrained to the rather low L/D ratios and, more particularly to the wide diversity of actual L and D values shown in Table 2. It should however be possible to overcome this restriction, if we can apply weighting effectively in this situation.

Let us therefore consider the application of weighting to the 'mean-squared' mode of operation. If the feature-dependent drive associated with a specific w, b is increased from 1 (due to row and column drives of +0·5 and −0·5 respectively) to $1+\delta$, then w 'white' elements and b 'black' elements are instantaneously driven to $(1+\delta)^2$ and the remaining $(W-w)$ 'whites' and $(B-b)$ 'blacks' are driven to δ^2. Hence the 'instantaneous mean-squared' contrast is then

$$\frac{[w(1+\delta)^2+(W-w)\delta^2]/W}{[b(1+\delta)^2+(B-b)\delta^2]/B}=\frac{\frac{w}{W}(1+2\delta)+\delta^2}{\frac{b}{B}(1+2\delta)+\delta^2}.$$

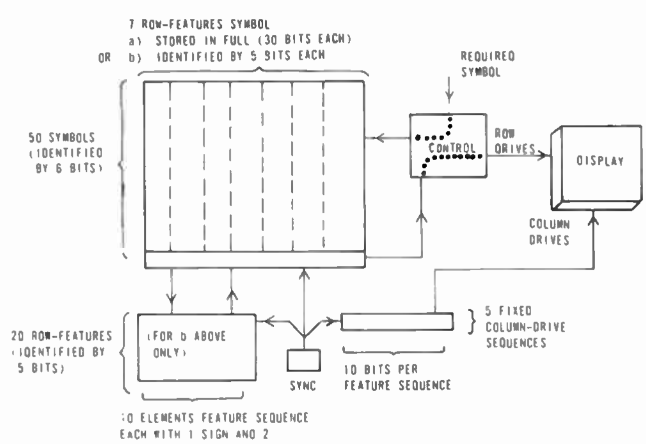


Fig. 6. Schematic of drive waveform generation and control.

Since, by definition, $w/W > b/B$, incrementing both equally can only degrade the contrast. Furthermore, the same conclusion applies if we decrease the drive (putting $\delta = -\varepsilon$). However, the unweighted mean-squared drive over the whole cycle is

$$\frac{[\sum_w N(w)]/W}{[\sum_w N(b)]/B}.$$

Hence it will pay to increase the weight of an above-average w/b by means of an increment δ, up to the point where its resulting decreased contrast denies us further benefit. Similarly, it will pay to decrease the weight of a below-average w/b by means of a decrement ε. (Note that here the increments δ and ε directly represent the change in the drive voltage—unlike Section 4.3, where weights are expressed in terms of their effect on the reflectance.)

Let the mean-squared light and dark levels for the total drive sequence be L_0 and D_0, in the absence of any weighting, and then let a particular combination w, b be incremented by δ. The resultant contrast is:

$$R = \frac{L_0 + k\left(2\dfrac{w}{W}\delta + \delta^2\right)}{D_0 + k\left(2\dfrac{b}{B}\delta + \delta^2\right)}$$

where

$$k \equiv N(w)/\sum N(w).$$

Differentiating this gives $\partial k/\partial\delta = 0$ for:

$$-\left(D_0\frac{w}{W} - L_0\frac{b}{B}\right) + \delta(L_0 - D_0) + k\delta^2\left(\frac{w}{W} - \frac{b}{B}\right) = 0$$

For $\delta \ll 1$ this gives

$$\delta \simeq \left(D_0\frac{w}{W} - L_0\frac{b}{B}\right)\Big/(L_0 - D_0).$$

More accurately

$$\delta = \frac{\left\{-(L_0 - D_0) + \sqrt{\left[(L_0 - D_0)^2 + 4k\left(\dfrac{w}{W} - \dfrac{b}{B}\right)\left(D_0\dfrac{w}{W} - L_0\dfrac{b}{B}\right)\right]}\right\}}{2k\left(\dfrac{w}{W} - \dfrac{b}{B}\right)}.$$

Note that these formulae give the correct magnitude and sign of the optimum δ, both for scaling-up the contribution from w, b combinations with better-than-average contrast, and for scaling down those worse than average. Direct application of even the simplified formula is normally sufficient, although exact calculation would entail some iteration, since each δ assignment modifies the L and D values to be used in computing the others. For $V = 5$, the poorest contrast, $R = 2{\cdot}33$, $(0{\cdot}7/0{\cdot}3)$ arises for $W = 2$ (see Table 2). Hence this analysis has been applied to improve this contrast by selective weighting to 2·7 (0·94/0·35). All the other W values for $V = 5$ have a larger actual or potential contrast, and have sufficient degrees of freedom that weights can be found, for them to match both the contrast ratio and the actual mean-squared drives (0·94 and 0·35) of $W = 2$. The conditions to be satisfied are

$$0{\cdot}94 = L_0 + \left[\sum N\left(2\frac{w}{W}\delta + \delta^2\right)\right]\Big/\sum N$$

$$0{\cdot}35 = D_0 + \left[\sum N\left(2\frac{b}{B}\delta + \delta^2\right)\right]\Big/\sum N$$

where L_0 and D_0 are the values of L and D in Table 2 (i.e. in the absence of any increments δ).

The increments thus found are listed in Table 6. It will be noted that a total of 12 increments (including zero) are used, and hence 4 bits would be required to identify the correct one, in the control signal (and memory). Some of these solutions are not unique, and hence some reduction in the diversity of drive levels may be feasible. However, this modest addition to the control circuits seems a small price to pay for the assurance of consistent, 'black' and 'white' levels, independent of the pattern being generated. The *consistency* would also be independent of the crystal material, but the actual black and white levels would be independent of the material only if the *ratio* of 2·7:1 in mean-squared drives is sufficient for switching between 'full black' and 'full white', and if the actual drive *levels* are correctly matched to the material used.

(Since the magnitude of the unweighted row drives is equal to that of the column drives, and is half the unweighted peak voltage, $\delta < -0{\cdot}5$ represents a reversal of the phase of the row drive; also $\delta > +1$ implies that the row drive is more than trebled and the peak-to-peak drive amplitude more than doubled. This however does not affect the validity of the formulae used.)

Table 6

Drive increments, for various values of W, with $V = 5$ to give constant mean-squared 'light' and 'dark' values of 0·94 and 0·35 respectively

W	w, b	δ
0	0, 3	−0·27
1	1, 1	+0·19
	1, 2	−0·21
2	2, 0	+0·80
	2, 1	+0·17
	1, 1	−0·20
3	3, 0	+1·08
	2, 0	0
	2, 1	−0·13
4	3, 0	−0.60
	2, 0	+0.58
5	3, 0	+0·24

5.3 Timing Considerations

In order to avoid an irritating flicker, due to the fluctuation of the control voltages, the luminance associated with each intersection must be smoothed by an appropriate integrating device. Hence each control voltage should go through identical (or at least perceptually equivalent) sequences in a period rather shorter than the smoothing 'time-constant' of the integrator. In the 'direct-response' mode, the response of the liquid crystal is fast enough to follow the voltage fluctuations. Hence smoothing is left to the human eye, and the complete cycle of control voltages must be kept rather shorter than the persistence of vision. In the 'root mean square' (r.m.s.) mode, the response time of the medium is rather longer than one full cycle of the control-voltage fluctuations, so that the smoothing is performed in the medium, and is independent of the characteristics of the observer.

In the 'direct-response' (d.r.) mode, the instantaneous voltage will generally switch each element to 'black' or 'white', and the perceived luminance depends on the relative proportions of the total time for which the instantaneous luminance is 'white' and 'black' respectively. If the eye is to average this, the repetiton period of the excitation waveforms should be shorter than the persistence of vision. However, generally the proportion of 'black' to 'white' clock cycles does not vary unduly over the total excitation period, and hence a somewhat longer excitation period—say 100 ms—would probably be perceptually acceptable. Since we have shown that the waveform period for a character display can be restricted to 10 clock cycles, this allows up to 10 ms per clock cycle. Hence any material response time constant less than, say, 5 ms should permit the 'direct-response' mode of operation.

In the r.m.s. mode, the effect of the control voltage is averaged over the repetition period of the excitation waveform (assumed to be less than the response time of the medium). Hence the resultant luminescence is steady, and is 'white' if this r.m.s. voltage is above an upper threshold, and 'black' if it is below a lower limit. This mode is less restrictive on timings. If the displayed pattern is to be able to change in not more than, say, 0·5 seconds, the time-constant of the material should not exceed 200 ms. This would permit a total waveform repetition period of up to 80 ms, and hence a clock cycle time of up to 8 ms. At the other limit, power consumption might discourage r.m.s. mode operation at clock frequencies exceeding 1 MHz, corresponding to a waveform period of 10 μs and a minimum material time constant of 20 μs, even if such materials were available.

These timing constraints are summarized in Table 7.

Whilst the r.m.s. mode is less critical with respect to response times and can potentially yield somewhat higher contrast than the direct response mode, it is more dependent on the precise characteristics of the material and on its operating conditions.

Table 7

Response time of material	R.M.S. mode	D.R. mode	Comment
> 200 ms	yes	—	Display changes too slow for some applications
< 200 ms > 5 ms	yes	—	—
< 5 ms > 20 μs	yes	yes	Direct response probably preferred
< 20 μs	—	yes	An unlikely response time for liquid crystals

In practice the response time is also a function of both the drive amplitude and of the immediately preceding drive condition, and operating conditions intermediate between the pure r.m.s. and direct-response modes are likely. Hence the weightings may require further (possibly empirical) adjustments to account for this. The small number of pattern-dependent sequences, and the short duration of these sequences, in the scheme here proposed make it relatively easy to find and implement these adjustments. (It may also be noted that the individual pattern-dependent sequences may not all have zero mean voltage. If the small d.c. components should prove undesirable, it might be worthwhile alternating between the basic set of row and column drive sequences and the complementary, sign-inverted set.)

5.4 Topological Problems for Arrays of such Dot-matrix Characters

The application of this form of drive to rows of characters poses two topological problems. The first is the requirement to connect to 7 row-drive terminals per character on, say, the back plane of the l.c.d. array. There should normally not be any major problem in bringing these connections out in the gaps between the characters. If all the characters are formed on a single plane, and all connections are to be made at the edge, this could however impose some restrictions on the inter-character spacing in the vertical and/or horizontal dimension. If the number of edge connections is therefore to be kept small, multiple row-drives may be multiplexed and then serial-to-parallel converted by shift registers. Placing these shift registers close to the relevant characters can then also reduce the demand for conductor space: a shift register along the left-hand edge of each character (or column of characters) reduces the preceding signal drives to one serial input per shift register. Several such shift registers can also time-share a common drive.

The second topological problem arises on the other face of the array, if we desire to use a single set of 5 common column-drive connections to the whole array without cross-over of conductor paths. This problem can clearly be avoided by replicating the set of five column-drive sequences once for each column of characters. Alternatively, alternate 'up' and 'down' drives can be used, as in Fig. 7(a), but in that case the horizontal order of element column drives will be reversed in alternate character columns, and so different row drives will be required for identical characters, depending on whether they are in even or odd-numbered character columns.

A second option is to provide an external 'fly-back', between the character columns as in Fig. 7(b); in that case all the character columns will behave identically. As a further alternative, the columns of characters can be arranged in one or more groups, with column-drive paths on the outside of the group. Figure 7(c) shows how this can be done with a single set of 5 feed points and terminations. Figure 7(d) shows how we can economize in connecting space at the expense of using two sets of 5 feed-points and terminations.

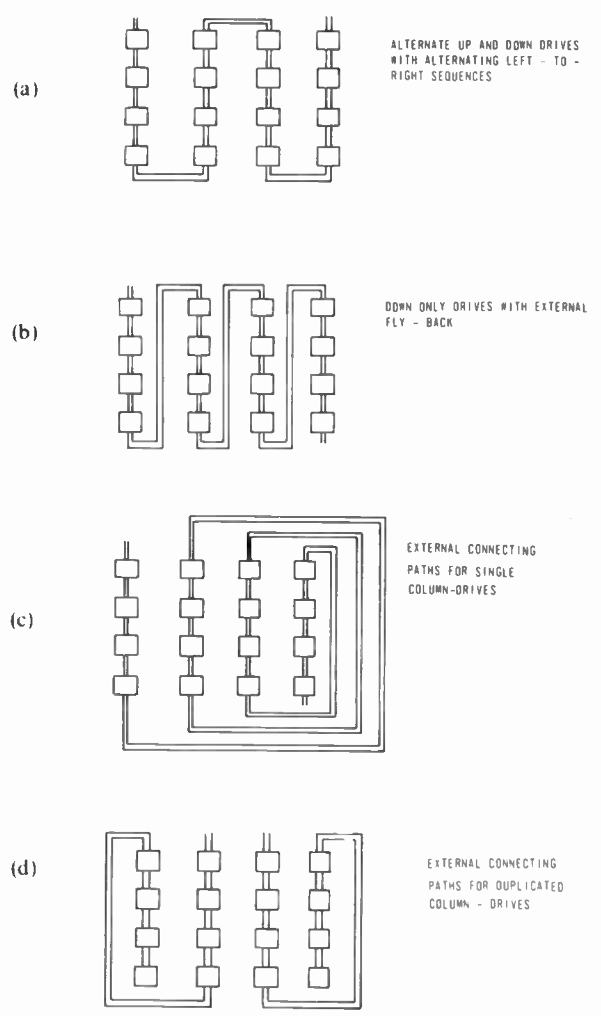


Fig. 7. Common drives to multiple columns of characters, whilst avoiding 'cross-overs'.

5.5 Waveform Displays

A second significant application of the technique is that of a waveform display, $y = F(x)$, where y is a single-valued function of x. This clearly needs a relatively large value of V (the range of y-values available), and there is no obvious benefit in this case in restricting V to odd values. As indicated in Table 2, for a 'black-on-white' display, the scheme should then give a true black level, and a reflectance of 0·5 (relative to 'full-white') for the 'white' remainder of each column. The vertical drive for each column-conductor would merely replicate the pre-determined row sequence of the row-conductor for the desired y, thus producing zero voltage difference at the junction as required. This scheme should work similarly with clock periods faster and slower than the relaxation time of the crystal.

For generating two such traces (or one double-valued function), the column drives have to revert to a form of 'majority voting'. With $V(V-1)/2$ double-valued and V single valued choices, these vertical features would probably be too many to store in advance (or to generate logically from stored functions of y_1 and $(y_1 - y_2)$, if that proved feasible in principle). Hence each vertical drive point would have to be fed with the row drives for its two y values. When these are both equal (both $+$ or both $-$), it would copy them. When they are unequal, we could choose $+$ and $-$ with equal probabilities. However, since these latter drives would contribute equally to the mean reflectivity of both 'black' and 'white' spots, it would clearly be preferable to apply zero drive, thus omitting them altogether *if crystal relaxation times short compared with the drive-element duration are achievable.*

For a triple trace, any such weighting would apply to all combinations of 2 '+' and 1 '−' (or vice versa) on the three wanted spots. In this case zero drive would reduce the reflectance of 'white' spots excessively, but half-amplitude drive may well be worthwhile; these conditions are listed in Table 8. (On this Table 'wtd' (= weighted) represents zero drive for the 1, 1 condition for two traces, and it represents half drive for the 2, 1 condition for three traces on the 'oscilloscope'.) Note the substantial improvements in contrast so obtained, particularly for black-on-white patterns. Thus acceptable performance would be attainable for double-valued and triple-valued functions (and even higher values of W may not be ruled out), given crystals of adequately short time-constant.

With such waveform displays V is almost bound to be far too large for the row-drive sequence to encompass all $V!/(V/2!)^2$ combinations of $V/2$ '+' and $V/2$ '−' drives in any sensible frame. Hence a subset of these would be chosen, which ensures that the V constituent row-line drives are mutually orthogonal, and this would allow the drive cycle to be reduced to a minimum duration of V clock periods. Even with this, however, the clock period for any visually acceptable frame duration would have to

Table 8

Light and dark levels (L and D) for waveform displays of two- and three-valued functions ($W = 2$ or $W = 3$). (For black-on-white 'W' should be interpreted as 'B'.) Crystal-response assumed to be immediate. wtd = weighted

		White-on-black		Black-on-white	
		un-wtd	wtd	un-wtd	wtd
$V = \infty$	L	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$
$W = 2$	D	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	0
$V = \infty$	L	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{5}{16}$
$W = 3$	D	$\frac{1}{2}$	$\frac{5}{16}$	$\frac{1}{4}$	$\frac{1}{8}$

be substantially shorter than currently attainable crystal relaxation times. Some extension of the frame period might be permissible, provided each non-selected spot will normally experience a representative mixture of drive values within the persistence of vision—even though the total frame time may be substantially longer. However, some noise or flicker effects in the 'white' background area would then probably be inevitable. This phenomenon could be much reduced or avoided altogether with a fast-cycle drive and the 'mean-squared' mode of operation.

As already indicated, 'mean-squared' operation does not alter the drive conditions or performance for a single-valued function. For a two-valued function, too, the optimum is still zero column drive when one wanted element has a '+' and the other a '−' row drive. This does however produce a net contribution to all elements, and hence some loss of contrast.

For a three-valued function, a slightly scaled-down column drive is warranted for a 2/1 split, but its potential benefit is rather marginal. See Table 9. (The optimum weight is $\sqrt{\frac{2}{3}}$, adequately approximated in the Table by 0·8.) It will be seen that black-on-white can, in all circumstances covered, give a mean-squared voltage ratio $\geqslant 2$. This is probably sufficient to give an acceptable performance with appropriate crystal materials.

Table 9

Light and dark levels (L and D) for waveform displays of two- and three-valued functions ($W = 2$ *or* $W = 3$). (For black-on-white 'W' should be interpreted as 'B'.) Mean-squared crystal response assumed. wtd = weighted

		White-on-black		Black-on-white	
		un-wtd	wtd	un-wtd	wtd
$V = \infty$	L	$\frac{3}{4}$	$\frac{5}{8}$	$\frac{1}{2}$	$\frac{3}{8}$
$W = 2$	D	$\frac{1}{2}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{8}$
$V = \infty$	L	$\frac{3}{4}$	0·58	$\frac{1}{2}$	0·38
$W = 3$	D	$\frac{1}{2}$	0·38	$\frac{1}{4}$	0·18

6 Conclusions

(1) A sandwich of

(i) a two-dimensional charge-coupled-device (c.c.d.) shift register array,
(ii) possibly a field-effect-transistor (f.e.t.) emitter-follower interface,
(iii) a two-dimensional liquid-crystal display (l.c.d.) array,

would provide a very versatile display device—including grey-scale reproduction—provided the technology were available.

(2) For more restricted but significant applications, crossed X and Y conductors can provide a very effective alternative, within the current 'state of the (technological) art'.

(3) The recommended solution involves an odd number of conductors in one plane, subject to a pre-determined sequence of drive patterns, each of quasi-equal numbers of '+' and '−' line-drive voltages.

(4) Each of the orthogonal conductors would then be driven by a binary sequence determined by the black-and-white pattern to be generated along it.

(5) In this simple form, better results are obtainable with black-on-white patterns than white-on-black if the pattern elements are in a minority.

(6) This and other undesirable pattern-sensitive characteristics of the scheme can be very substantially ameliorated by using, say, four distinct amplitude levels in the pattern-dependent drive.

(7) In particular, this technique, allied to the *row-feature* concept, permits very effective 7×5 dot-matrix characters to be generated in a relatively simple and inexpensive manner.

(8) There are reasonable topologies for using such a row-feature drive in lieu of the more obvious column-feature approach.

(9) The basic technique is also well adapted to the relatively simple generation of single-valued waveform displays.

(10) It should indeed be adaptable for the display of at least two- and probably three-valued waveforms. These however will entail some loss of contrast and brilliance, together with increased complexity of the drive-generation for the vertical conductors.

(11) The drive-scheme here derived should permit the relevant waveforms for dot-matrix character displays to be generated at a rate slow enough for the crystals to respond.

(12) For oscilloscope-type waveform displays, however, the scheme proposed probably does not simplify the drive cycle sufficiently for currently available crystals to respond to it other than in a 'mean-squared, slow-response' manner.

(13) The choice between the direct response (integrated within the response time of the observer's

eye) or the mean-squared, slow-response, mode of operation (integrated within the relaxation time of the crystal) will have to be made with a particular application in mind. The former appears well matched to character displays and has the advantages of low-speed operation, simplicity of design and construction and robustness to changes in the operating conditions of the crystal. The latter, providing that appropriate weighting can be applied to provide sufficient mean-squared voltage contrast for the operating conditions prevailing, can provide a better visual performance.

7 References

1 Benjamin, R., Improvements in or relating to matrix addressing of opto-electronic displays. Patent Application Number 7,934,975, 7th October 1979.

2 Clark, M. G., Shanks, I. A. and Patterson, N. J., 'General theory of matrix addressing liquid crystal displays', SID Digest 1979, Paper 13.1.

3 Nehring, J. and Kmetz, A. R., 'Ultimate limits for matrix addressing of rms-responding liquid crystal displays', *IEEE Trans. on Electron Devices*, **ED-26**, 1979, in the press.

4 Shanks, I., Method and Apparatus for Matrix Addressing Opto-Electric Displays. Patent Application Number 30897/78.

Manuscript received by the Institution on 4th October 1979
(Paper No. 1927/CC 321)

Standard Frequency Transmissions

(Communication from the National Physical Laboratory)

Relative Phase Readings in Microseconds
NPL—Station
(Readings at 1500 UTC)

JANUARY 1980	MSF 60 kHz	GBR 16 kHz	Droitwich 200 kHz
1	−3·9	7·2	36·2
2	−4·2	8·4	36·0
3	−4·2	6/0	35·8
4	−4·0	8·5	35·6
5	−4·0	8·2	35·4
6	−4·2	6·7	35·3
7	−4·2	8·6	35·0
8	−4·3	6·5	34·7
9	−4·2	8·5	34·5
10	−4·5	6·2	34·2
11	−4·7	8·4	33·8
12	−4·6	5·5	33·4
13	−4·6	7·1	33·1
14	−4·4	9·0	32·9
15	−4·6	7·5	32·7
16	−4·4	7·0	32·4
17	−4·3	7·9	32·2
18	−4·3	7·7	31·9
19	−4·3	8·1	31·7
20	−4·3	8·0	31·5
21	−4·2	7·0	31·3
22	−4·2	4·7	31·2
23	−4·0	7·8	31·1
24	−4·2	6·9	30·9
25	−4·2	7·2	30·7
26	−4·4	7·0	30·5
27	−4·5	6·3	30·2
28	−4·4	7·2	30·0
29	−4·4	9·7	29·8
30	−4·5	9·4	29·7
31	−4·4	8·2	29·6

Notes: (a) Relative to UTC scale (UTC_{NPL}–Station) = +10 at 1500 UTC, 1st January 1977.
(b) The convention followed is that a decrease in phase reading represents an increase in frequency.
(c) Phase differences may be converted to frequency differences by using the fact that 1 μs represents a frequency change of 1 part in 10^{11} per day.

UDC 681.3.04: 681.327.63/4
Indexing Terms: Recording, magnetic, Data storage, digital, Coding

The choice of a recording code

N. D. MACKINTOSH, B.Sc., Ph.D.*

Based on a paper presented at the IERE Conference on Video and Data Recording held in Southampton from 24th to 27th July 1979

SUMMARY

A detailed comparison of recording codes, using superposition, is used to show that there is little to choose between the popular ones, as regards maximum achievable data density. It is shown that relatively few general classifications are necessary to encompass all the codes, and that, within each classification, it is often possible to say which code is optimum. A guide to the selection of an efficient code for some common conditions of use is presented.

* *Racal Recorders Limited, Hythe, Hampshire, England; now with Burroughs Corporation, 5411 North Lindero Canyon Road, Westlake Village, CA 91361, USA.*

1 Introduction

The choice of a recording code is of fundamental importance in the design of any moving magnetic digital storage system. Many papers have recently appeared extolling the virtues of yet another new code, and proving how superior it is to all other codes. Such papers are generally followed by another (biased) one proving that the new code is not as good as one of the old ones. As a result, anyone faced with the choice of a 'best' code in any particular situation can be forgiven for thinking it to be a more difficult task even than the selection of the 'best' microprocessor for a given application!

In an effort to alleviate this situation, a detailed, unbiased comparison of the vast majority of possible codes has been undertaken, using a superposition program. The superposition technique itself and the choice of the 'basic pulse' (isolated reversal response) are described in a companion paper.[1] To make any comparison between different recording codes valid, it is necessary to ensure that the conditions of use are the same for each code. To do this strictly would be to ensure that all codes were tested on the same mechanical disk unit, with fixed amounts of noise present, using exactly the same writing and reading electronics, and therefore with exactly the same mechanical and electrical tolerances. In practice, however, with a comparison which involves simulation rather than actual experimentation, it is sufficient to ensure that a standard basic pulse response is used throughout, with a fixed signal-to-noise ratio and a fixed level of error, pertaining to all tolerances and inaccuracies throughout the system. This last point is very important, and the errors due to these factors can conveniently be compounded into one figure, called real time error (*RTE*), and introduced into the analysis as clocking inaccuracies. *RTE* is thus typically composed of many factors, such as crosstalk, incomplete erasure effects, print-through, plating noise, component tolerances, phase-locked loop errors, and even an allowance for margin, if desired.

The results of this analysis of codes are presented for four different values of *RTE*, to allow a system designer to interpolate between the specific figures given to relate to the actual *RTE* anticipated in his own system. The first *RTE* allowance is $RTE = 0$, resulting in the theoretical maximum possible packing factor, *PF* (data density, normalized to PW_{50}, the half-amplitude basic pulse width). The second and third allowances are $RTE = 8\%\ PW_{50}$ and $RTE = 16\%\ PW_{50}$, being representative of clocking errors occurring in practice. The fourth allowance is an attempt to reflect the self-clocking ability of certain codes. Codes with regular transitions (having, therefore, a low value of maximum inter-reversal time, $\hat{IRT}$) are penalized in terms of flux-reversal density, but produce a more stable clock waveform. The fourth *RTE* allowance is therefore $RTE \propto \hat{IRT}$. Obviously the exact nature of the proportionality is difficult to define, but

0033–7722/80/040177+17 $1.50/0

based on measurements of an actual disk system, the following equation was determined:

$$RTE = (6{\cdot}7 + 1{\cdot}3 \times \hat{IRT})\%\ PW_{50}.$$

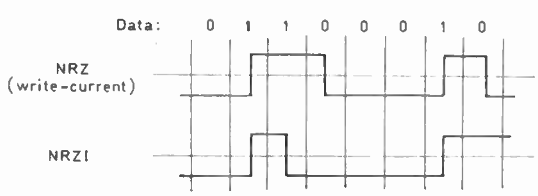
Note that for a code with an $\hat{IRT}$ of $1{\cdot}0 \times BP$ (bit period), this conveniently yields an RTE of 8% PW_{50}, as in the second RTE allowance.

It would be unfair to use the same detection system for all of the codes, and so, in each case, an optimum detection system for the code under consideration is determined. The final answer is then presented as the maximum normalized data frequency achievable by the code, i.e. maximum packing factor. Note that $PF = PW_{50}/BP$, but for codes where data frequency does not equal reversal frequency, a distinction will be made between data bit period (DBP) and reversal bit period (RBP).

2 NRZ/NRZI (Non-Return-to-Zero and Non-Return-to-Zero Modified)

2.1 Coding Rules

(a) NRZ: Change direction of saturation at m.b.t. (mid-bit time) only if present bit $\neq$ previous bit.
(b) NRZI: Change at m.b.t. only if bit is a '1'.



As both codes are able to produce the same waveforms, and differ only in such matters as error propagation, parity bit generation and, perhaps, ease of understanding, the subsequent analysis of packing density limit will be confined to just one of the codes, NRZI.

2.2 Rectify and Clip Detection System

2.2.1 *Circuit description*

In this detection system, the signal is first amplified, then full-wave rectified, and then clipped to remove baseline noise, which would otherwise be a problem later on, in the squaring process. The noise present at the input to the linear amplifier will have come from several sources and a certain amount of this can be filtered out by correct choice of the frequency response of the linear amplifier, but a problem arises when the noise has components at frequencies lower than the maximum significant frequency in the 'basic pulse'. Attempted filtration of these components will result in integration of the data waveform, having the effect of increasing the PW_{50} of the basic pulse. Since the slimness of the latter is all-important for efficient use of the recording unit, it is vital that no attempt is made to filter these components.

Experiments on a typical recording unit produced the result that, with as much filtering as possible employed, concurrent with no integration of the data, the signal-to-noise ratio is typically 20 : 1 at the output of the linear amplifier.

For the determination of the absolute limit of the code and detection method, no margins are allowed, and so the clip-level is set to the peak noise level, normalized, of 0·05. Note that this assumes zero baseline shift.

The next requirement is to detect the peaks in the waveform by converting them into zero-crossovers, using differentiation. This involves the subtraction of a delayed version of the signal from itself, and it has been shown[1,2] that an optimum value of delay is $\sim 0{\cdot}3 \times PW_{50}$.

The signal is next squared, using a comparator, to detect the zero-crossovers (z.x.o.s).

2.2.2 *Packing density limit*

The packing density achievable using this detection system may be limited by timing considerations, amplitude considerations, or a mixture of the two. Considering first the timing; the worst-case pattern for peak-shift is two 1's, and this shift is shown in Fig. 1 as a function of frequency. It can be seen that the theoretical ideal frequency limit for NRZI with zero RTE is 2·33, where the peak-shift equals 50%, the read resolution for this code. Further, by plotting another graph, of worst-case peak-shift (as above) +8% PW_{50} RTE, it is seen that the frequency limit for NRZI with a system RTE of 8% PW_{50} is 1·92. A similar calculation shows the timing limit for $RTE = 16\%PW_{50}$ to be at $PF = 1{\cdot}62$. Considering now the amplitude limit, this is where a peak in the read signal does not exceed the clip-level. The worst-case amplitude pattern at any packing density is three 1's, and is plotted in Fig. 2. It falls to the clip-level of 0·05 at $PF = 1{\cdot}88$.

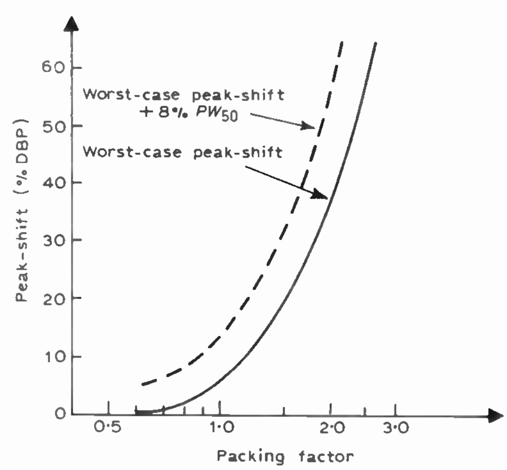


Fig. 1. Worst-case peak-shift for NRZI.

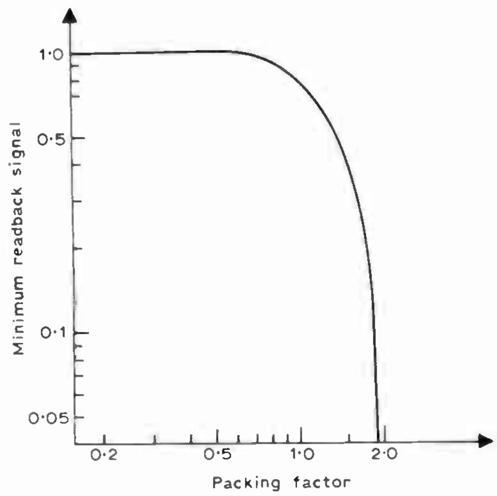


Fig. 2. Minimum read-back signal for NRZI.

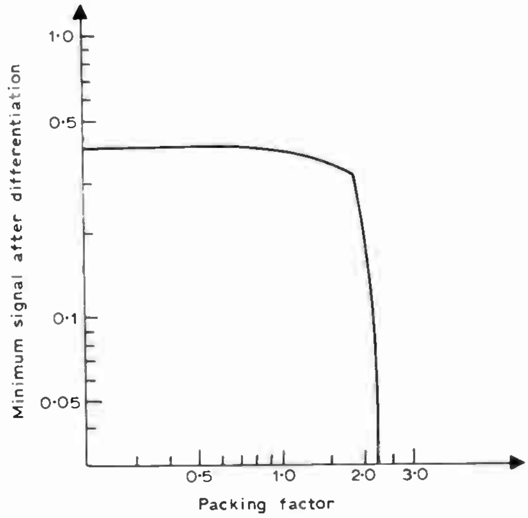


Fig. 3. Minimum signal after differentiation for NRZI.

2.2.3 *Summary*

The limiting frequency for each *RTE* allowance can now be summarized:

$RTE = 0$: Although the timing limit for this case is 2·33, an amplitude limit occurs first at 1·88.

$RTE = 8\%\ PW_{50}$: Again, although the timing does not limit performance until 1·92, breakdown occurs at 1·88, because of amplitude.

$RTE = 16\%\ PW_{50}$: Timing causes the breakdown in this case, at 1·62.

$RTE \propto I\hat{R}T$: Since $I\hat{R}T = \infty$, and $RTE = (6{\cdot}7 + 1{\cdot}3 \times IRT)\%\ PW_{50}$, then operation is impossible under these conditions, and the limiting frequency is zero.

2.3 Differentiate and Square Detection System

2.3.1 *Circuit description*

An alternative version of the previous detection method is to eliminate the clipping (and, therefore, the need to rectify), and to differentiate the amplified read-signal directly. Because of the large amount of noise present after differentiation, the squaring must then be biased considerably. This may be achieved by hysteresis round a comparator, so that the reference level is alternately +0·1 and −0·1 (i.e. the noise amplitude, allowing for doubling of the noise in the differentiator).

2.3.2 *Packing density limit*

The main disadvantage of this method is that, in the worst-case derivative waveform, there is an appreciable time delay between a zero cross-over and the 0·1 or −0·1 level, resulting in a timing error. Neglecting this for the moment, the amplitude limit will be where the derivative does not exceed the bias level of ±0·1. This worst-case derivative amplitude is plotted against frequency in Fig. 3.

The portion before the 'break-point' at 1·87 is due to the trailing edge of any pattern of consecutive 1's, whilst that after is due to the centre portion of a four 1's pattern. This latter breakdown features in many codes and so is shown in Fig. 4 in detail, at the amplitude limit for this method of 2·14. The low slope of the read-back waveform between the two inner 1's produces a very small peak in the derivative.

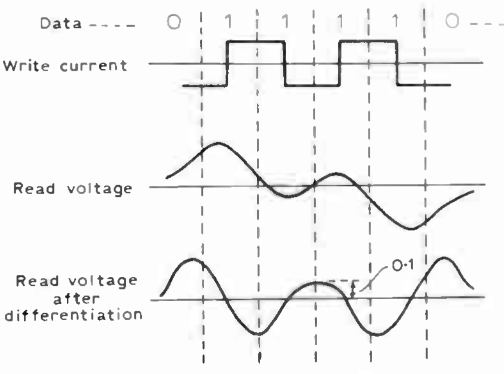


Fig. 4. Amplitude limit due to 'four-1's' pattern.

$RTE = 0$: On top of the normal timing loss due to worst-case peak-shift, a further timing loss is introduced by the biased squaring. At the amplitude limit of 2·14, this latter loss is $\pm 17\%\ BP$, whilst the peak-shift on two 1's is $\pm 43{\cdot}5\%\ BP$. The total is therefore much greater than the permissible $50\%\ BP$. The timing limit is found to be at $PF = 2{\cdot}00$, where worst-case peak-shift is $37{\cdot}5\%\ BP$, and the biased-squaring loss is $12{\cdot}5\%\ BP$, a total of $50\%\ BP$. So, for $RTE = 0$, the limit is 2·00.

$RTE = 8\%\ PW_{50}$: At $PF = 1{\cdot}85$, worst-case peak-shift $= 31{\cdot}5\%\ BP$, $RTE = 14{\cdot}8\%\ BP$, and squaring loss $= 3{\cdot}7\%\ BP$, a total of $50\%\ BP$. Thus the limit is 1·85.

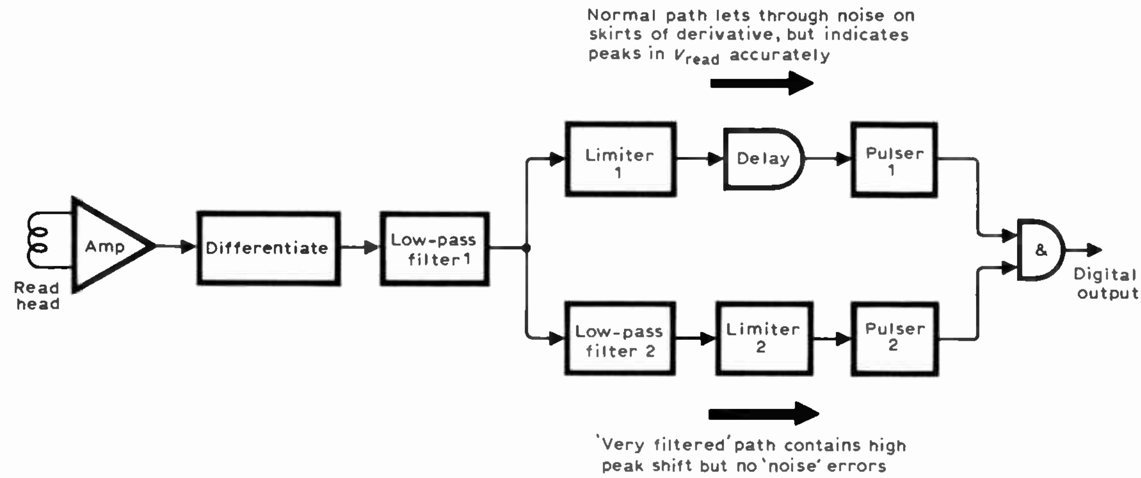


Fig. 5. Gated crossover detection method.

$RTE = 16\% \, PW_{50}$: A similar calculation shows the limit to be 1·60.

$RTE \propto \hat{IRT}$: As in the previous detection method, the limit here is zero.

2.3.3 *Summary*

The four packing factor limits for NRZI with this detection system are:

$$\left.\begin{array}{ll} RTE = 0: & \hat{PF} = 2{\cdot}00 \\ RTE = 8\% \, PW_{50}: & \hat{PF} = 1{\cdot}85 \\ RTE = 16\% \, PW_{50}: & \hat{PF} = 1{\cdot}60 \\ RTE \propto \hat{IRT}: & \hat{PF} = 0{\cdot}0 \end{array}\right\} \text{All due to timing}$$

2.4 Gated Cross-over Detection System

2.4.1 *Circuit description*

The main disadvantage of the previous method is the timing loss introduced by the biased squaring. In the gated cross-over method, the zero cross-overs are used to indicate the peak positions in the read waveform, so that there is no timing loss, and the '0·1 cross-overs' are used to discriminate against noise on the baseline.

The method was first proposed by Dunstan and Whitehouse[3,4] and a similar system also appears to have been used by Tamura *et al.*[5] As shown in Fig. 5 two paths are formed after differentiation, one 'very filtered' path which contains high peak-shift, but no errors due to noise, and one unfiltered path which lets baseline noise through but indicates peaks in the readback waveform accurately. The two are then combined to give a noise-free, accurate, digital output.

2.4.2 *Packing density limit*

$RTE = 0$: The timing limit for this case is where the worst-case peak shift = 50% BP, i.e. at $PF = 2{\cdot}33$. However, an amplitude limit occurs where a signal peak after differentiation = 0·1, at $PF = 2{\cdot}14$. The limit is therefore at 2·14.

$RTE = 8\% \, PW_{50}$: The timing limit for this case is where (worst-case peak shift + jitter) = 50% BP, i.e. at 1·92. As the amplitude limit occurs at a higher frequency, 1·92 is the limit in this case.

$RTE = 16\% \, PW_{50}$: The limit is 1·62 due to timing.

$RTE \propto \hat{IRT}$: As before, operation is not possible in this case.

2.4.3 *Summary*

The four limits for this detection system are:

$$\begin{array}{lll} RTE = 0: & \hat{PF} = 2{\cdot}14 & \text{due to amplitude} \\ RTE = 8\% \, PW_{50}: & \hat{PF} = 1{\cdot}92 & \\ RTE = 16\% \, PW_{50}: & \hat{PF} = 1{\cdot}62 & \text{due to timing} \\ RTE \propto \hat{IRT}: & \hat{PF} = 0 & \end{array}$$

2.5 Pattern-adaptive Writing

2.5.1 *Basic technique*

This technique involves the modification of the timing of the write-current for certain specific data patterns. It can be used to trade-off amplitude against timing, or vice versa, depending on which is causing the most problems. As an example, in NRZI it has been shown that the peak-shift produced by a 'two-1's' pattern can become intolerable at high frequencies. Thus, in Fig. 6 it can be seen that by writing the two transitions at m.b.t., as is normal, the peaks in the read waveform are shifted, say 50% of a bit period away from the correct m.b.t. positions. If, however, this two-1's pattern is recognized before writing, and the two transitions are written closer than they would normally be, the peaks in the read waveform are found to be shifted less, relative to m.b.t., than in the normal method. The penalty, as can be seen, is a reduction in peak amplitude. Conversely, by writing the two transitions further apart than normal, an increase in amplitude can be obtained, at the expense of greater peak-shift. It should be noted that pattern

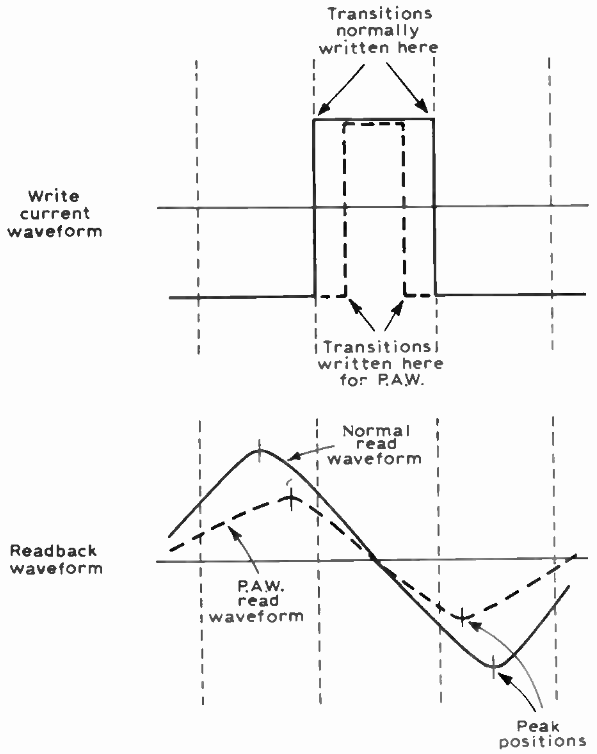
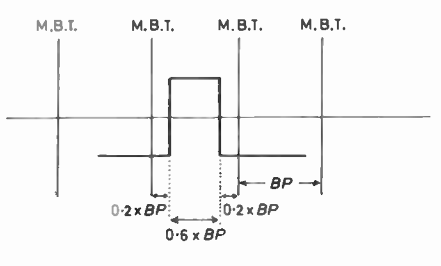


Fig. 6. Pattern adaptive write technique.

adaptive write (p.a.w.) techniques can be applied to any pattern, not just two-1's.

The notation used to describe the extent to which p.a.w. is being used will be as follows: if a transition is written a distance of $n\%$ BP away from its normal position, and in such a direction as to produce less effective peak-shift than it would in the normal position, then $x\%$ p.a.w. is being used, where $x = (50-n) \times 2$. For the case where p.a.w. is being used to produce *more* peak-shift than normal n will be negative, and x will be greater than 100. For example, in the case shown below, 60% p.a.w. is being used.



It can be seen, therefore, that for the common case of 'two-1's', $x\%$ conveniently represents the separation of the two 1's, in terms of bit-periods. Note that 100% p.a.w. is equivalent to *no* p.a.w.

2.5.2 *Application to NRZI*

Consider the case of 8% PW_{50} RTE with the gated cross-over method. This breaks down at 1·92 because of excessive peak-shift in the two-1's pattern. It is necessary to know how much p.a.w. to apply to this pattern to take the achievable packing density up to the next limit, at 2·14, where the amplitude of the derivative of four-1's causes breakdown.

Allowable peak-shift at 2·14

$$= 50\% BP - RTE$$
$$= 50\%\ BP - 8\% \times 2{\cdot}14 \times BP$$
$$= 32{\cdot}9\%\ BP$$

or, in terms of PW_{50},

allowable peak separation for two-1's

$$= BP + 2 \times 0{\cdot}329 \times BP$$
$$= 1{\cdot}66 \times BP$$
$$= (1{\cdot}66 - 2{\cdot}14) \times PW_{50}$$
$$= 0{\cdot}78 \times PW_{50}.$$

The amount of p.a.w. necessary to produce this separation is most conveniently determined by the use of Fig. 7(a) which is a plot of written-transition separation against read-back-peak separation, for the two-1's pattern. Note that at low packing densities, i.e. written separation greater than approximately $1{\cdot}8 \times PW_{50}$, no interaction takes place between the two readback pulses;

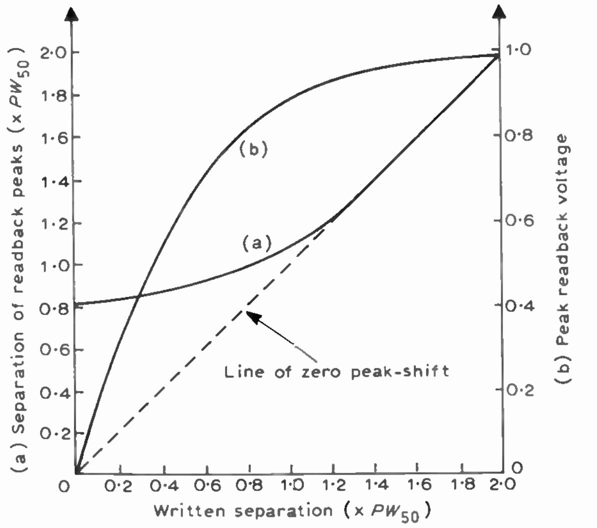


Fig. 7. Read-back separation and peak voltage for two isolated flux-reversals.

there is therefore no peak-shift, and so read-back separation = written separation. As the two transitions are written progressively closer, however, intersymbol interference produces peak-shift, and so read-back separation > written separation. In the limit, as the written separation tends to zero, the read-back separation tends to a finite limit of $0{\cdot}81 \times PW_{50}$.

To see why this arises, we can consider that by placing a second 'isolated' pulse close to the first one, as we are doing, the effect is to differentiate the first pulse. As the separation decreases, more accurate differentiation is achieved, and so at the limit of zero separation, the ideal derivative is produced. This will have peaks at the points of maximum slope on the original pulse. These points occur at a distance of $0{\cdot}405 \times PW_{50}$ either side of the peak, thus agreeing with the asymptote of the graph.

The graph shows, therefore, that no amount of p.a.w. will allow NRZI to work at 2·14 with a RTE of 8% PW_{50}, because a two-1's read-back separation of $0{\cdot}78 \times PW_{50}$ cannot be achieved. A theoretical timing limit occurs at 2·04 when 0% p.a.w. is used, though of course the peak amplitude of the read-back waveform is zero as shown in Fig. 7(b), which plots peak amplitude against peak separation, and operation under such circumstances is impossible.

One further point is that the peak-shift on the end 1's of three-1's, four-1's, five-1's, or, indeed, any number of 'isolated' ones, is nearly as bad as the absolute worst-case two-1's pattern. This means that if p.a.w. was used in a system, the criterion used would probably be: apply p.a.w. to any 1 if it has three or more 0's on one side of it, and one or more 1's on the other side. To determine the practical limit for p.a.w. using the gated crossover detection system, therefore, involves the simultaneous study of both amplitude and timing effects for all patterns, at all frequencies, and with all extents of p.a.w.

2.5.3 *Summary*

The result of such an analysis is that p.a.w. can only produce a practical packing density increase from 1·92 to 2·00, for 8% PW_{50} RTE. This is an increase of only 4%. For 16% PW_{50} RTE, the improvement is even less. The conclusion is, therefore, that p.a.w. is not worthwhile as a means of increasing packing density for NRZI.

2.6 Summary of NRZ/NRZI Systems

RTE	Rectify and clip limit	Differentiate and square limit	Gated cross-over limit
0	1·88	2·00	2·14
8% PW_{50}	1·88	1·85	1·92
16% PW_{50}	1·62	1·60	1·62

3 Enhanced NRZI

It should be noted that enhancement can also be applied to NRZ, but, as before, NRZI only will be discussed.

3.1 Coding Rules

Type A. Code as in NRZI, but after every n bits include one compulsory '1'.

Type B. Code as in NRZI, but after every n bits include one odd-parity bit.†

† It can be easily shown that, for NRZ, a further stipulation is that n must be odd, to ensure a finite $I\hat{R}T$.

3.2 Description

In this code, it is convenient to make a distinction between recorded-bit-frequency, or flux-reversal frequency, and average data frequency, since for every $n+1$ bits recorded, only n of them are data bits. Thus, we have data frequency $= n/(n+1) \times$ reversal frequency, and since data bit period $DBP = PW_{50}$/data frequency, and reversal bit period $RBP = PW_{50}$/reversal frequency, then $DBP = RBP \times (n+1)/n$.

The purpose of this code is to utilize the good qualities of NRZI, i.e. ±50% read resolution, absence of 'double frequency' components, and ease of coding/decoding, whilst removing its main disadvantage, its inability to provide self-clocking. With enhanced NRZI (ENRZI), by appropriate choice of n, any required degree of self-clocking can be obtained, using the following formulae:

for type A:

$$I\hat{R}T = (n+1) \times RBP = n \times DBP$$

for type B:

$$I\hat{R}T = (2n+1) \times RBP = [(2n+1)n/(n+1)] \times DBP.$$

For all values of n, but $n = 1$, both types of ENRZI exhibit effectively the same worst-case patterns as NRZI, and will break down, therefore, at the same reversal frequency as NRZI, and, more importantly, at the same reversal and data frequency as each other. Thus, given any value of n other than 1, both types of ENRZI will break down at the same data frequency, whilst type A will exhibit an $I\hat{R}T$ typically half that of type B, and will therefore be better at self-clocking. Detailed analysis of worst-case patterns shows that for $n = 1$, type A is also superior in terms of achievable data frequency, as well as in terms of self-clocking ability.

It can be seen that although type B assists error detection by virtue of the fact that the compulsory bit is a parity bit, it is never better than type A as regards achievable packing density, and is always worse at providing self-clocking. For this reason, type A is considered superior to type B, and the former alone will now be analysed in detail.

3.3 Packing Density Limit

3.3.1 $n \geq 2$

For $n \geq 2$, ENRZI exhibits effectively the same worst-case patterns as NRZI, and, therefore, if the gated crossover detection method is used, the packing density limits can be easily calculated by applying the dilution factor to the corresponding NRZI limit.

(a) $RTE = 0$. For NRZI the limit here is 2·14. Thus for ENRZI the limits are:

n and $I\hat{R}T$ ($\times DBP$)	2	3	4	5	6	7	8	9	10	x	∞
$\hat{PF}$	1·43	1·61	1·71	1·78	1·83	1·87	1·90	1·92	1·94	$\frac{2{\cdot}14x}{(x+1)}$	2·14

(b) $RTE = 8\%\ PW_{50}$. For NRZI the limit here is 1·92, and so for ENRZI the limits are:

n and $\hat{IRT}$ ($\times DBP$)	2	3	4	5	6	7	8	9	10	x	∞
$\hat{PF}$	1·28	1·44	1·54	1·60	1·65	1·68	1·71	1·73	1·75	$\frac{1{\cdot}92x}{x+1}$	1·92

(c) $RTE = 16\%\ PW_{50}$. For NRZI, the limit here is 1·62, and so for ENRZI the limits are:

n and $\hat{IRT}$ ($\times DBP$)	2	3	4	5	6	7	8	9	10	x	∞
$\hat{PF}$	1·08	1·22	1·30	1·35	1·39	1·42	1·44	1·46	1·47	$\frac{1{\cdot}62x}{x+1}$	1·62

(d) $RTE \propto \hat{IRT}$. For this case, it is not possible to work straight from the NRZI figure, as there is no exact parallel between the two codes. For each value of n, therefore, the procedure for determining the data frequency limit is as follows:

e.g. for $n = 3$:

$$RTE = (6{\cdot}7 + 1{\cdot}3 \times \hat{IRT}/DBP)\%\ PW_{50}$$
$$= (6{\cdot}7 + 3{\cdot}9)\%\ PW_{50}$$

therefore

$$RTE = 10{\cdot}6\%\ PW_{50}.$$

The timing limit occurs where (RTE + worst-case peak-shift) = 50% RBP. From the graph of worst-case NRZI peak-shift, this occurs at 1·81, where (peak-shift $+RTE$) = 30·8% RBP + 19·2% RBP = 50% RBP.

As this occurs before the amplitude limit at 2·14, the reversal-frequency limit is 1·81.

Therefore data frequency limit

$$= n/(n+1) \times \text{reversal frequency limit}$$
$$= (\tfrac{3}{4}) \times 1{\cdot}81.$$

Hence for $n = 3$, data frequency limit = 1·36.

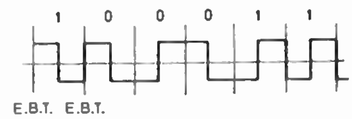
By a similar process, the corresponding limits for other values of n can be found, resulting in:

n and $\hat{IRT}$ ($\times DBP$)	2	3	4	5	6	7	8	9	10	20	34	∞
$\hat{PF}$	1·23	1·36	1·42	1·44	1·44	1·47	1·40	1·40	1·37	1·16	0	0
RTE (%PW_{50})	9·3	10·6	11·9	13·2	14·5	15·8	17·1	18·4	19·7	32·7	50·1	∞

It is interesting to note that, for this RTE case, maximum packing density is achieved by using $n = 7$, which is the version most often used in practice.

3.3.2 $n = 1$

ENRZI with $n = 1$ produces the well-known frequency modulation code (FM), the usual definition of which is: always change at m.b.t., but if the bit is a '1', change at e.b.t. also, e.g.:



Another very similar code is phase modulation code (PM), and other names for the two codes include 'Manchester', Biphase and FSK, but, as with NRZI and NRZ, there is no need to consider both PM and FM in detail, as they produce exactly the same waveform sets. Several detection techniques were applied to FM, including gated crossover, polarity strobing at m.b.t., polarity strobing at e.b.t., and derivative polarity strobing at t.q.b.t. (three-quarter bit-time).

The exceedingly simple detection method of strobing polarity at m.b.t. proves to be superior to all others considered, yielding the following figures:

(i) for $RTE = 0$: $\hat{PF} = 1{\cdot}17$;
(ii) for $RTE = 8\%\ PW_{50}$ or $RTE \propto \hat{IRT}$: $\hat{PF} = 1{\cdot}15$;
(iii) for $RTE = 16\%\ PW_{50}$: $\hat{PF} = 1{\cdot}09$.

4 Modified NRZI

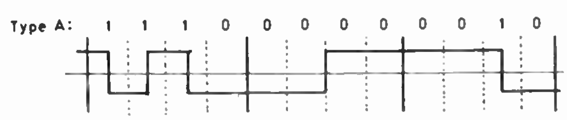
As before, modification can also be applied to NRZ, but NRZI only will be considered.

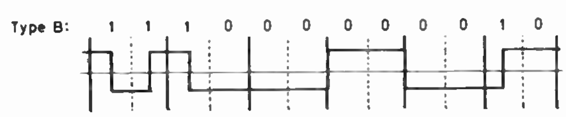
4.1 Coding Rules

Type A. Code groups of $2n$ bits at a time, and code each group as in NRZI, except code a group of all 0's as a change at mid-group time.

Type B. Code groups of n bits at a time, as in NRZI, but code a group of all 0's followed by the same as a change on the junction of the two groups.

For example $n = 2$:





4.2 Description

Like ENRZI, modified NRZI (MNRZI) is an attempt to produce a self-clocking version of NRZI, in this case by breaking up long runs of zero's with an occasional e.b.t. (end bit time) pulse. It might appear at first that by inserting e.b.t. pulses in isolated areas, in this manner, a self-clocking code will be produced without affecting the normal operation of NRZI, and so the code will work to

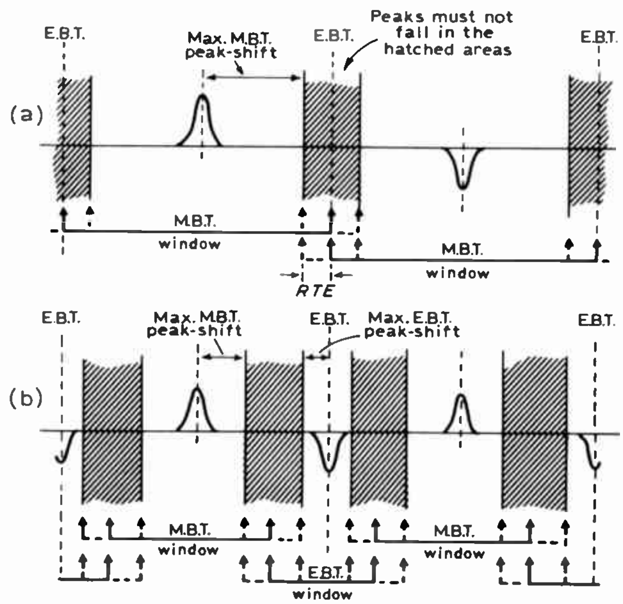


Fig. 8. Reduction of margin by adding e.b.t. transitions to NRZI.

the same limits as NRZI. Unfortunately, this is not the case, as can be seen with reference to Fig. 8. (a) signifies NRZI with a m.b.t. window which has non-zero jitter (RTE). The window has a width of BP, and the jitter, or RTE, is effectively shared between adjacent windows, so that the peak-shift allowable on the m.b.t. pulse = ($50\% \, BP - RTE$). In (b), which depicts any code that contains both m.b.t. and e.b.t. pulses, the jitter is no longer shared by adjacent m.b.t. windows, but by adjacent m.b.t. and e.b.t. windows (whether or not there is an actual e.b.t. window). It is clear that the allowable peak-shift on the m.b.t. pulse is now ($50\% \, BP - 2 \times RTE -$ max. e.b.t. peak-shift). So even if the e.b.t. shift is negligible, the allowable m.b.t. peak-shift is RTE less than for 'straight' NRZI.

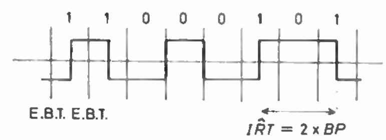
From this it can be expected that for very small RTE values, MNRZI will closely resemble NRZI in achievable packing density, the exact proximity of the two depending on the extent of self-clocking required. As the RTE figure increases, however, the MNRZI performance will rapidly decrease, showing MNRZI to be not a worthwhile code for high RTE values. This will be borne out in the ensuing detailed analyses, in which the various MNRZI codes are referred to as MNRZI_{xn}, where x is either 'A' or 'B', indicating the type, and n is as defined in the coding rules.

4.3 MNRZI_{B1}

4.3.1 *Coding rules*

This code is the well-known Modified Frequency Modulation, described by Padalino[6] (attributed by Padalino to Pouliart[7]), but also described by Woo.[8] It is also known as Miller Code, Delay Modulation (DM), and, in a slightly different form, Modified Phase Modulation (MPM).

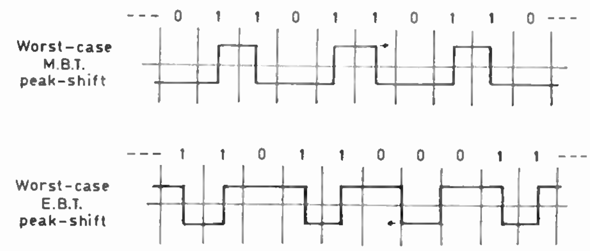
The coding rules are: change at m.b.t. for a 1, and change at e.b.t. between two 0's. For example:



4.3.2 *Gated crossover detection*

As shown in Section 4.2, a timing limit occurs for MNRZI when (worst-case m.b.t. peak-shift $+ 2 \times RTE$ + worst-case e.b.t. peak-shift) = $50\% BP$.

For Miller code the two worst-case patterns are shown below:



The graphs of peak-shift against packing factor for the two patterns are shown in Fig. 9, from which the timing limit for any value of RTE can be found:

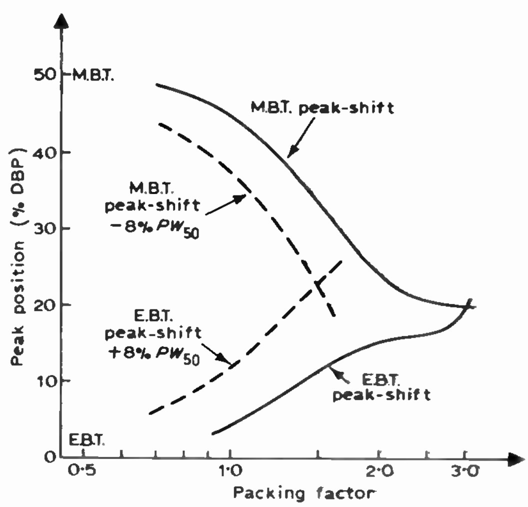
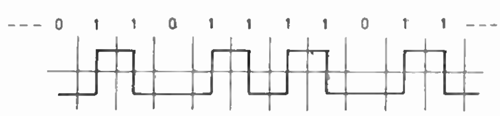


Fig. 9. Worst-case peak-shift for Miller code.

(a) $RTE = 0$: The intercept of the two curves gives directly the timing limit for zero RTE as 2·95. An amplitude limit will occur when the peak amplitude of the derivative is equal to the noise level, i.e. 0·1. This occurs at 2·14 for the four-1's-type pattern:



Thus, for zero RTE, DM breaks down at 2·14, due to amplitude.

(b) $RTE = 8\%\ PW_{50}$: The timing limit for this value of RTE is obtained by replotting the two curves to include the RTE, as shown. The intercept of the two, at 1·47, is the timing limit. As the amplitude limit is unchanged at 2·14, then for $8\%\ PW_{50}\ RTE$, DM breaks down at 1·47.

(c) $RTE = 16\%\ PW_{50}$: By replotting the curves (though this is not shown, for clarity), the timing limit is found to be 1·12. As this is below the amplitude limit, breakdown is at 1·12.

(d) $RTE \propto I\hat{R}T$: For DM,

$$I\hat{R}T = 2 \times BP.$$

$$RTE = (6{\cdot}7 + 2 \times 1{\cdot}2)\%\ PW_{50} = 9{\cdot}3\%\ PW_{50}.$$

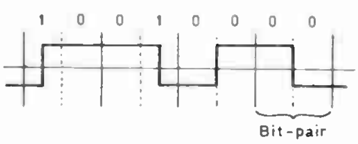
The timing limit is found to be 1·4 and as this is below the amplitude limit DM breaks down at 1·4, for the fourth RTE case.

4.4 MNRZI$_{A1}$

4.4.1 *Coding rules*

Code bits in pairs, and code as in NRZI, except code a pair of 0's as a change at mid-pair-time (m.p.t.).

This yields an $I\hat{R}T$ of $3 \times BP$, as in the following pattern:



4.4.2 *Packing density limit*

Analysis of the worst case peak-shift and amplitude patterns yields the following results:

$RTE = 0$:	$P\hat{F} = 1{\cdot}97$
$RTE = 8\%\ PW_{50}$:	$P\hat{F} = 1{\cdot}47$
$RTE = 16\%\ PW_{50}$:	$P\hat{F} = 1{\cdot}15$
$RTE \propto I\hat{R}T = 10{\cdot}6\% \times PW_{50}$:	$P\hat{F} = 1{\cdot}36$.

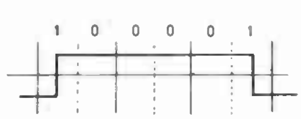
All the limits are due to timing, as the amplitude limit (with the gated crossover detection method) is at 2·14.

4.5 MNRZI$_{B2}$

4.5.1 *Coding rules*

Code bits in pairs, and code pairs as in NRZI, except code a pair of 0's followed by the same as a change on the boundary of the two pairs.

The maximum inter-reversal time is $5 \times BP$, as in the pattern:



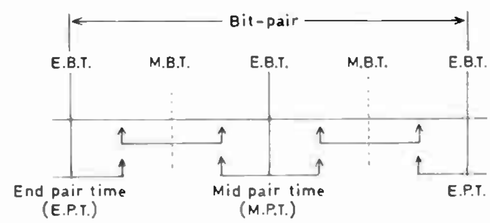
4.5.2 *Packing density limit*

$RTE = 0$:	$P\hat{F} = 2{\cdot}06$
$RTE = 8\%\ PW_{50}$:	$P\hat{F} = 1{\cdot}57$
$RTE = 16\%\ PW_{50}$:	$P\hat{F} = 1{\cdot}22$
$RTE \propto I\hat{R}T = 13{\cdot}2\%\ PW_{50}$:	$P\hat{F} = 1{\cdot}30$.

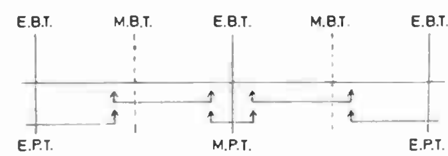
All limits are again due to timing, as the amplitude breakdown (gated crossover detection) is 2·14.

4.5.3 *Window modification*

In this code, for the first time, a slightly different method of arranging the m.b.t. and e.b.t. windows is possible. In the normal method, the windows are exactly the same for each bit, regardless of its position in the pair, thus:



However, if (i) the worst-case peak-shift from m.b.t. → e.p.t. is not the same as that from m.b.t. → m.p.t., and (ii) e.p.t. → m.b.t. is different to m.p.t. → m.b.t., then the windows can be different for each bit in the pair, e.g.:

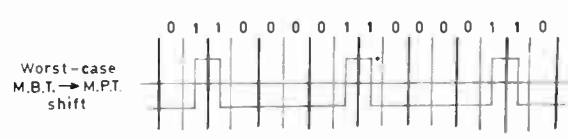


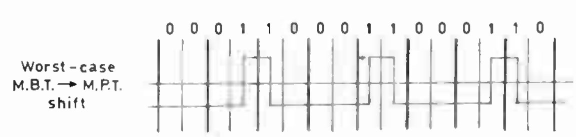
The structure just drawn is the one that could be applied to this code, because:

(a) There is no m.p.t. pulse, so condition (ii) is met, and

(b) The worst-case m.b.t. → m.p.t. shift is greater than that from m.b.t. → e.p.t., so condition (i) is satisfied.

Unfortunately, however, the two shifts mentioned in (b) above are very close to each other, as might be guessed from the patterns producing them:





Thus, the increase in performance possible in this case is found to be only 2% (up from 1·57 to 1·60) (for the $8\%\ PW_{50}\ RTE$ case), and is not considered worthwhile.

4.6 Summary

The performance of all the MNRZI codes possible is summarized in Fig. 10. The results verify the original postulate that MNRZI would not be worthwhile for high RTE values. Even for the 8% PW_{50} RTE allowance, the best MNRZI code will only work up to 1·65, where (two-1's peak-shift$+2\times RTE$) = 50% BP. If no e.b.t. pulses have to be allowed for, as in NRZI, the limit is 1·92 where (two-1's peak-shift$+1\times RTE$) = 50% BP.

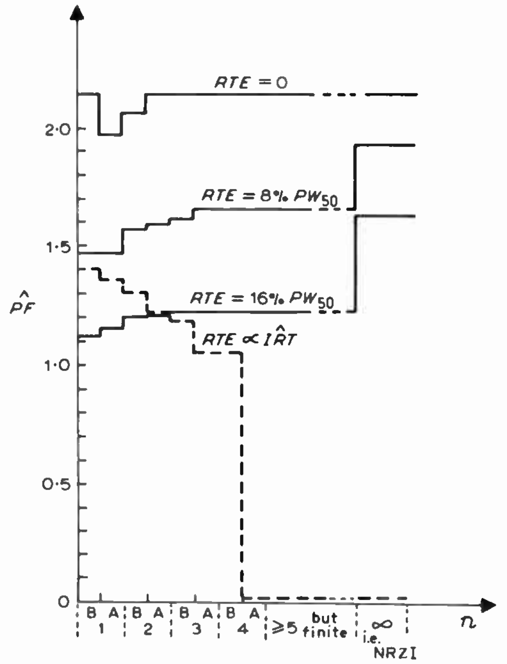


Fig. 10. MNRZI performance summary.

One important conclusion is that if $RTE \propto I\hat{R}T$, no trade-off is necessary, as $MNRZI_{B1}$ (Miller) provides both the highest achievable frequency and the lowest $I\hat{R}T$, showing why Miller code is a very popular code in practice.

5 Group Codes

Codes such as NRZ, NRZI, PM and FM operate on the principle of 'one symbol for one bit'. There is no reason, however, why codes should not be constructed whereby groups of bits are coded with unique patterns. Such codes are called group codes, and indeed, MNRZI could be viewed as one.

A similar class of codes is adaptive codes, whereby groups are viewed, as before, but the waveform symbol depends not only on the group presently being coded, but also on previous and subsequent groups. The distinction between the two classes is vague, however, and both will herein be referred to as group codes.

Franaszek's paper[9] is an attempt to provide a means of producing optimum group codes, given the constraints of minimum and maximum inter-reversal times.

The symbols used by Franaszek are:

N = number of equal subdivisions per bit.
= number of possible flux-reversal positions per bit (e.g. for FM, $N = 2$).
d = minimum number of empty flux-reversal positions between two flux-reversals.
k = maximum number of empty flux-reversal positions between two flux-reversals; thus

$$I\check{R}T = [(d+1)/N]\times BP$$

and

$$I\hat{R}T = [(k+1)/N]\times BP.$$

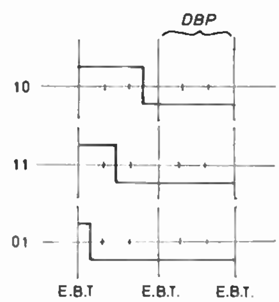
M = maximum number of bits required for coding at any one time (e.g. for FM, $M = 1$; but for Watson code, $M = 2$).

This notation will be used for all the group codes presented here.

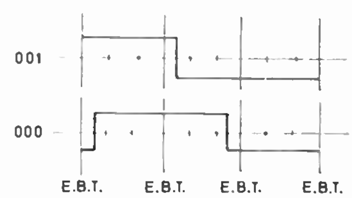
5.2 Group Code (4, 9)

5.2.1 *Coding rules*

This code is given by Franaszek as an example of a run-length limited code. Coding commences by the viewing of the first two data bits. If these are anything other than a pair of 0's, the waveforms shown below are applied.



If, however, the pair is 00, the third bit is viewed also, and the three bits together are then coded thus:



The parameter values are thus:

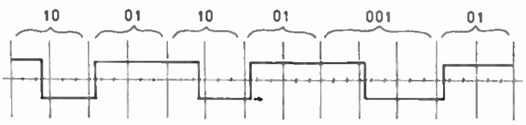
$$N = 3,\quad d = 4,\quad k = 9,\quad M = 3$$

and

$$I\hat{R}T = 3{\cdot}33\times BP,\quad I\check{R}T = 1{\cdot}67\times BP.$$

5.2.2 *Packing density limit*

The read resolution for this code is only $\pm 16{\cdot}7\%$ BP. The worst-case peak-shift pattern is:



(a) $RTE = 0$: The peak-shift equals 16·7% at $PF = 1{\cdot}70$. There is no problem with the amplitude with any detection system at this frequency, so breakdown is at 1·70.
(b) $RTE = 8\%\ PW_{50}$: The timing limit here is at 1·30.
(c) $RTE = 16\%\ PW_{50}$: This gives a timing limit of 1·06.
(d) $RTE \propto I\hat{R}T$: Since $I\hat{R}T = 3{\cdot}33 \times BP$, $RTE = 11{\cdot}0\%\ PW_{50}$, and this breaks down at 1·20.

5.3 Rice Code

5.3.1 *Coding rules*

The first six bits are viewed initially. If they are '010101', they are coded as shown in Fig. 11; otherwise the last two bits are returned to the input stream for now. If the remaining bits commence with '11', then all four are coded, as shown. Otherwise, the last two bits are returned to the input stream, and the remaining two bits are coded as shown. The process is then repreated, starting with the next six bits in the input stream (including returned bits).

The parameters are:

$$N = 1\tfrac{1}{2},\quad d = 1,\quad k = 11,\quad M = 6$$

and

$$I\hat{R}T = 8 \times BP,\quad I\check{R}T = 1{\cdot}33 \times BP.$$

The read resolution is $\pm 33{\cdot}3\%\ BP$.

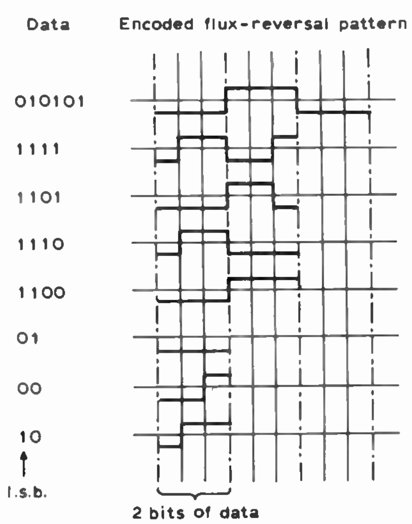


Fig. 11. Coding rules for Rice code.

5.3.2 *Packing density limit*

The worst-case peak-shift pattern is shown in Fig. 12(a). An amplitude limit occurs for a four-1's-type pattern, shown in Fig. 12(b), at 2·84, assuming a gated-crossover detection system.

(a) $RTE = 0$: The timing limit is where worst-case peak-shift $= 33\tfrac{1}{3}\%\ BP$, at $PF = 2{\cdot}23$. This occurs before the amplitude limit, and therefore breakdown is at 2·23.

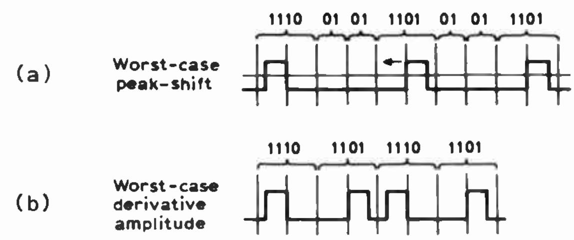


Fig. 12. Rice code worst-case patterns.

(b) $RTE = 8\%\ PW_{50}$: The timing limit for this case, and therefore breakdown, occurs at 1·77.
(c) $RTE = 16\%\ PW_{50}$: This gives a timing limit at 1·47.
(d) $RTE \propto I\hat{R}T$: Since $I\hat{R}T = 8 \times BP$, $RTE = (6{\cdot}7 + 1{\cdot}3 \times 8)\%\ PW_{50} = 17{\cdot}1\%\ PW_{50}$. The timing limit for this is at 1·41.

5.3.3 *Summary*

This code, used by Digital Development Operation, is obviously of considerable interest as it provides a performance comparable with that of NRZI, whilst allowing some degree of self-clocking. Because of the large difference between the amplitude and timing limits, it might appear prudent to use p.a.w. However, because of the complexity of the code, it is difficult to decide where exactly to apply the p.a.w. An alternative is pulse-slimming, which may well improve its performance even further.

5.4 Gabor Code

5.4.1 *Coding rules*

This code, proposed by Gabor,[10] is a very complicated adaptive code. Bits are viewed in groups of two, and each double bit-period is subdivided into three parts, each of which is a possible flux reversal position.

The notation used is as follows:

	Data bits	Code bits
Preceding bit pair	B_{1p} B_{2p}	P_{1p} P_{2p} P_{3p}
Present bit pair	B_1 B_2	P_1 P_2 P_3
Following bit pair	B_{1f} B_{2f}	P_{1f} P_{2f} P_{3f}

The code is constructed so as to obey:

$$I\hat{R}T = 4/3 \times BP;\quad I\check{R}T = 2/3 \times BP.$$

The formal encoding rules are:

$$P_1 = \overline{P_{3p}} + B_1 + \overline{B}_2 B_{1f}$$

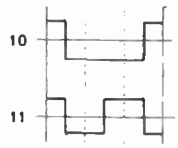
$$P_2 = \overline{P_{3p}}\,\overline{B}_1 + B_2$$

$$P_3 = \overline{P_{3p}} + B_1 + B_2$$

This is most easily visualized as:

If the bit pair is '10' *or* '11'

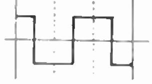
then code as follows:



else if the bit pair is '01';

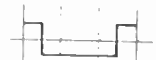
then if P_{3p} code as: →

or if $\overline{P_{3p}}$ code as: →



'00' is the awkward case, and is coded as:

if $\overline{P_{3p}}$: →



if P_{3p} *and* B_{1f}: →

if P_{3p} *and* $\overline{B_{1f}}$: →

The decoding rules are:

$$B_1 = [\overline{P_{3p}} P_{1p} + P_{3p} P_1 P_3]; \quad B_2 = [P_2 P_3].$$

5.4.2 *Packing density limit*

The read resolution for this code is again 33·3% BP, and analysis of worst-case patterns yields the following results:

(a) $RTE = 0$: A peak-shift of 33·3% is never reached, so breakdown is due to amplitude at 1·45.
(b) $RTE = 8\%\ PW_{50}$: A timing limit now occurs at 1·75 but breakdown is still due to amplitude at 1·45.
(c) $RTE = 16\%\ PW_{50}$: Timing now causes breakdown at 1·15.
(d) $RTE \propto I\hat{R}T$: Since $I\hat{R}T = 1{\cdot}33 \times BP$, $RTE = 8{\cdot}43\%\ PW_{50}$, and breakdown is at 1·45 due to amplitude, as the timing limit is at 1·7.

In such a complex code as this, p.a.w. would be difficult to apply.

5.5 Octal Coded Binary

5.5.1 *Coding rules*

In this code, data are coded in groups of three. The waveform sets for octal-coded-binary (OCB) code are shown in Fig. 13. As in all the other codes presented here, inverses of waveforms are non-distinct, and are used to ensure that $I\hat{R}T = 4{\cdot}2 \times BP$, subject to $I\check{R}T = 1{\cdot}2 \times BP$.

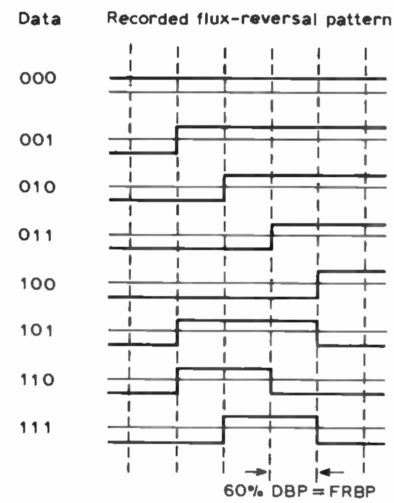


Fig. 13. Waveforms for octal-coded-binary.

5.5.2 *Packing density limit*

(a) $RTE = 0$: The worst-case peak-shift curve shows that 30% DBP (the read-resolution for this code) occurs at 2·03. As the amplitude limit is at 2·56, breakdown is at 2·03 due to timing.
(b) $RTE = 8\%\ PW_{50}$: This breakdown occurs at 1·62, again due to timing.
(c) $RTE = 16\%\ PW_{50}$: Breakdown here is at 1·30.
(d) $RTE \propto I\hat{R}T$: Since $I\hat{R}T = 4{\cdot}2 \times DBP$, $RTE = 12{\cdot}16\%\ PW_{50}$. With this RTE figure, the new (timing) breakdown is at 1·45.

5.5.3 *Pattern adaptive write*

Because the amplitude limit is significantly higher than the timing limit, it might be expected that p.a.w. could be profitably used. However, because the timing limit is quite high anyway, very little reduction in peak-shift can be obtained, even with large amounts of p.a.w. The increase obtainable is in fact only $\sim 5\%$.

5.6 'GCR'

5.6.1 *Coding rules*

In this code, also known as '4/5 code' and analysed also by Tamura *et al.*,[5] four data bits are represented by a five-bit pattern. The constraints placed upon the code are that

$$I\hat{R}T = 2{\cdot}4 \times DBP\ (= 3 \times FRBP)$$

and

$$I\check{R}T = 0{\cdot}8 \times DBP\ (= 1 \times FRBP).$$

From the 32 possible combinations of 5 bits, 15 can be eliminated because of these constraints, leaving 17, from which one can be discarded to produce the 16 unique patterns required. This one can then be used as a special pattern, for checking or error detection, as it obeys the constraints and is therefore detectable.

5.6.2 *Packing density limit*

Considering the gated crossover type, amplitude breakdown will occur at $PF = 1{\cdot}74$ for a four-1's type pattern. The read resolution of this code is $50\%\ FRBP = 40\%\ DBP$, so:

(a) $RTE = 0$: The timing limit occurs at $PF = 2{\cdot}12$. As the amplitude limit occurs earlier, however, the code breaks down at 1·74.
(b) $RTE = 8\%\ PW_{50}$: The modified graph shows that the new timing limit is at 1·57, and as this is below the amplitude limit, breakdown therefore occurs at 1·57.
(c) $RTE = 16\%\ PW_{50}$. Breakdown here is at 1·31, against due to timing.
(d) $RTE \propto I\hat{R}T$. Since

$$I\hat{R}T = 3 \times FRBP = 3 \times 0{\cdot}8 \times DBP$$
$$= 2{\cdot}4 \times DBP,$$

then

$$RTE = (6{\cdot}7 + 1{\cdot}3 \times 2{\cdot}4)\%\ PW_{50}$$
$$= 9{\cdot}82\%\ PW_{50}.$$

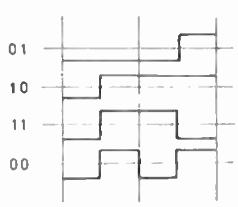
This timing limit occurs at $PF = 1{\cdot}50$, and this is where breakdown occurs.

5.7 Watson Code

5.7.1 *Coding rules*

Code bits in pairs, and code as in NRZI, except code a pair of 0's as a change at both m.b.t.s *and* a change at mid-pair-time (m.p.t.).

Thus the four possible patterns for a pair are:



In this code, proposed by Watson,[11] runs of zeros are broken up by the use of a unique flux-reversal pattern, which the originator hopes will be still easily identifiable at high packing densities.

5.7.2 *Packing density limit*

Using a slightly modified gated crossover detection method, an amplitude limit occurs at 1·30 due to the inherent three-1's-type pattern of the '00' bit pair.

(a) $RTE = 0$: A timing limit occurs at 2·33, where the peak-shift = $50\%\ BP$. However, amplitude causes an earlier limit at 1·30.
(b) $RTE = 8\%\ PW_{50}$: Although the timing limit is at 1·92, amplitude again causes breakdown at 1·30.
(c) $RTE = 16\%\ PW_{50}$: Timing limit = 1·65, but amplitude limit = 1·30.
(d) $RTE \propto I\hat{R}T$: $I\hat{R}T = 3 \times BP$, $RTE = 10{\cdot}6\%\ PW_{50}$, giving a timing limit at 1·82, but amplitude fails at 1·30.

5.7.3 *P.A.W.*

This code provides an opportunity for 'reverse' p.a.w., because of its early amplitude breakdown. If the two outer flux-reversals in the '00' pattern are written further away from the centre one, the timing margin will be reduced, but the amplitude of the centre peak will increase. Note that this is not strictly p.a.w., but merely a modification of the coding rules. For each separate *RTE* allowance, the amount of 'p.a.w.' can be optimized to ensure that the timing limit and the amplitude limit occur simultaneously. With the gated crossover detection method, this yields:

(a) $RTE = 0$: By using -40% p.a.w., on the outer peaks of the '00' pattern only, the frequency limit becomes 1·60, where the shift on one of these peaks takes it to e.b.t., and, simultaneously, the four-1's type pattern breaks down on amplitude.
(b) $RTE = 8\%\ PW_{50}$: The optimum in this case is -70% p.a.w., causing simultaneous timing and amplitude breakdowns at 1·50.
(c) $RTE = 16\%\ PW_{50}$: In this case, the limit is at 1·36, with -90% p.a.w.
(d) $RTE \propto I\hat{R}T$: With -80% p.a.w., this breakdown is at 1·42.

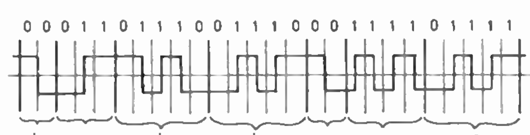
5.8 Miller² Code

5.8.1 *Coding rules*

The bit stream to be encoded is broken into sequences of three types:

(a) Any number of consecutive ones.
(b) Two zeros separated by either no ones, or an odd number of ones.
(c) One zero followed by an even number of ones (terminated by a zero not counted as part of the sequence).

Sequences type (a) and (b) are coded as in normal Miller code. Sequences type (c) have the transition corresponding to the final '1' inhibited, e.g.:

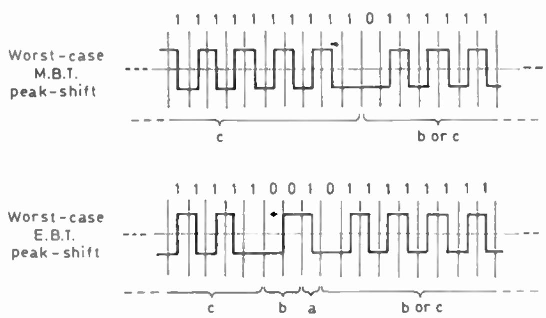


This code, invented by J. W. Miller,[12] is a modification of the original Miller code (sometimes attributed to A. Miller[13]), in such a way as to remove its d.c. content. The d.c. content of a code manifests itself as a baseline shift of either the read-back waveform, or its derivative. This can cause errors in a detection system which is particularly sensitive to amplitude variations, but can be overcome to a large extent by d.c. restoration circuitry. The alternative is to use a code such as Miller,[2] with zero d.c. content, but the disadvantage that results is a greater $\hat{IRT}$ ($3 \times DBP$), yielding higher peak-shift and requiring more sophisticated clocking circuitry.

Note that this code is also known as M^2, or M^2FM, but is not the same as MMFM, which is very similar to $MNRZI_{B2}$.

5.8.2 *Packing density limit*

The worst case m.b.t. and e.b.t. peak-shift occurs for the following patterns:



These patterns are almost identical to the worst-case patterns for $MNRZI_{A1}$, for which $\hat{IRT}$ is also $3 \times DBP$. Performance limits are thus (from the corresponding $MNRZI_{A1}$ figures):

$$\left.\begin{array}{ll} RTE = 0: & \hat{PF} = 1{\cdot}97 \\ RTE = 8\%\ PW_{50}: & \hat{PF} = 1{\cdot}47 \\ RTE = 16\%\ PW_{50}: & \hat{PF} = 1{\cdot}5 \\ RTE \propto \hat{IRT}: & \hat{PF} = 1{\cdot}36 \end{array}\right\} \text{all due to timing.}$$

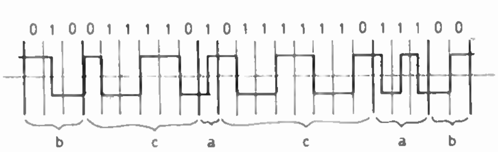
5.9 Zero Modulation Code

5.9.1 *Coding rules*

The bit stream to be encoded is broken into sequences of three types:

(a) Any number of consecutive ones.
(b) Two zeros separated by either no ones, or an odd number of ones.
(c) Two zeros separated by an even number of ones.

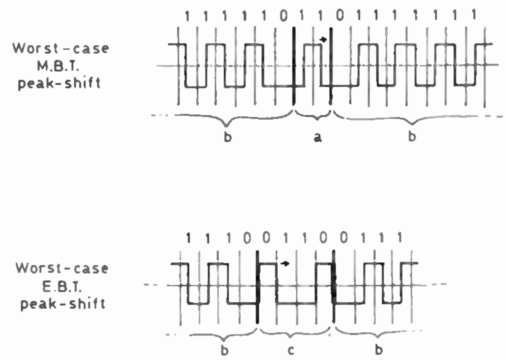
Sequences type (a) and (b) are coded as in normal Miller code. In sequences type (c), ZM encodes the zeros in the Miller manner, but the ones are encoded as though they were zeros but with alternate transitions deleted:



This code, invented by Patel,[14] is another attempt at improving Miller code by removing its d.c. content. It has the same disadvantages as Miller,[2] though $\hat{IRT}$ is only $2 \times DBP$.

5.9.2 *Packing density limit*

The worst-case peak-shift patterns are as follows:



Pattern (a) is the same as the worst-case m.b.t. shift for Miller code, whilst pattern (b) has greater peak-shift than the corresponding pattern for Miller. Analysis of these patterns yields the following packing density limits:

$$\left.\begin{array}{ll} RTE = 0: & \hat{PF} = 1{\cdot}94 \\ RTE = 8\%\ PW_{50}: & \hat{PF} = 1{\cdot}40 \\ RTE = 16\%\ PW_{50}: & \hat{PF} = 1{\cdot}11 \\ RTE \propto \hat{IRT}: & \hat{PR} = 1{\cdot}32 \end{array}\right\} \text{all due to timing.}$$

This analysis is based on ideal zero modulation code, which is very difficult to implement since complete sequences have to be modified, requiring infinite look-forward and look-back memories (unlike Miller,[2] where changes are introduced only at the end of sequences). Patel recognized this difficulty, and suggested that the memory could be reduced by blocking the data into groups of bits followed by a parity. Naturally, this dilutes the data to an extent dependent on the length of a block, but performance can never be better than the figures given above for the infinite memory.

5.10 3PM Code

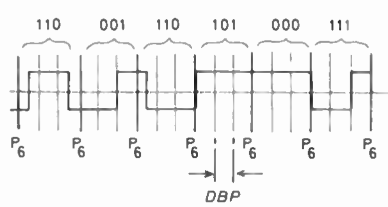
5.10.1 *Coding rules*

The data stream is split into groups of three bits, which are then encoded into a six-bit word for recording. The code is adaptive, in that the bit pattern for the word currently being encoded depends on the previous and the following word. The code has been designed to produce $\check{IRT} = 1{\cdot}5 \times DBP$ and $\hat{IRT} = 6 \times DBP$. The coding is represented by:

Data	Previous P_5	Following P_1	Recorded transitions P_1	P_2	P_3	P_4	P_5	P_6
000	×	0	0	0	0	0	1	0
		1	0	0	0	0	0	1
001	×	×	0	0	0	1	0	0
010	×	×	0	1	0	0	0	0
011	×	0	0	1	0	0	1	0
		1	0	1	0	0	0	1
100	×	×	0	0	1	0	0	0
101	0	×	1	0	0	0	0	0
	1		0	0	0	0	0	0
110	0	0	1	0	0	0	1	0
	1		0	0	0	0	1	0
	0	1	1	0	0	0	0	1
	1		0	0	0	0	0	1
111	0	×	1	0	0	1	0	0
	1		0	0	0	1	0	0

(× = Don't care)

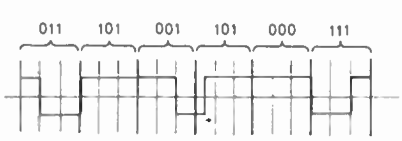
e.g.



This code, presented by Jacoby,[15] is essentially a standard type of adaptive block code, but is unusual in achieving an $I\check{R}T$ of $1{\cdot}5 \times DBP$. This lessens the worst-case peak shift, but because $I\hat{R}T = 6 \times DBP$, the code is difficult to clock accurately.

5.10.2 *Packing density limit*

The worst-case peak-shift pattern is:



This is plotted in Fig. 14, from which the following packing factor limits can be obtained (since the detection window is $\pm 25\%$ DBP):

$$RTE = 0: \quad \hat{PF} = 2{\cdot}13$$
$$RTE = 8\%\ PW_{50}: \quad \hat{PF} = 1{\cdot}67$$
$$RTE = 16\%\ PW_{50}: \quad \hat{PF} = 1{\cdot}29$$
$$RTE \propto I\hat{R}T: \quad \hat{PF} = 1{\cdot}36$$

All these limits are due to timing, as amplitude is no problem for any detection system.

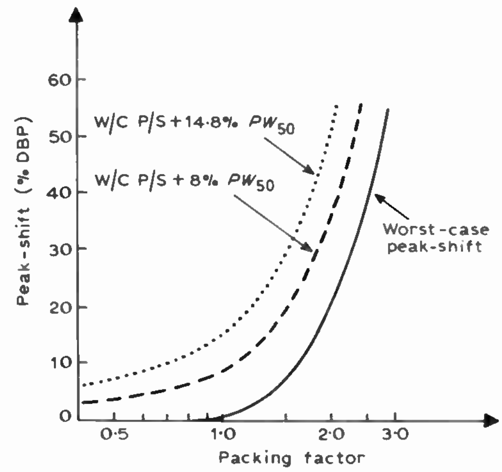


Fig. 14. Worst-case peak shift for 3PM code.

6 Summary of Code Performances

The performance of the best of the codes is summarized in Figs. 15 and 16. Figure 15 shows the maximum packing factor achievable by each code for the three fixed values of RTE. For $RTE = 0$, NRZI is surprisingly beaten for first place by the relatively unknown Rice code, because of its very good $I\check{R}T$ of $1{\cdot}33 \times DBP$, and yet surprisingly large window of $33\frac{1}{3}\% \times DBP$. There is little to choose between the top five codes, in fact, in this RTE category.

For $RTE = 8\%\ PW_{50}$, NRZI is slightly better than its nearest rivals—Rice code, $ENRZI_7$, 3PM and OCB. Again, for $RTE = 16\%\ PW_{50}$, NRZI achieves maximum performance, with Rice code and $ENRZI_7$ fairly close behind.

Perhaps the most useful indication of the performance of the codes is for $RTE \propto I\hat{R}T$, shown in Fig. 16. The striking result is that, apart from PM (which is only used when packing density is unimportant), all the other codes can achieve packing densities which are within $\pm 6\%$ of each other! Additionally, the three most recent codes, i.e. Miller,[2] 3PM and ZM, are of below average ability, showing that it may be more prudent to spend money on d.c. restoration circuitry rather than complicated encode and decode electronics.

In support of the theoretical work presented here, and in particular the general implications of Fig. 16, consider the results presented by several other authors:

Tamura *et al.*[5] compared GCR, FM and Miller, and found packing density limits of GCR : FM : Miller = 1·08 : 0·73 : 1·0.

Huber[16] found MFM : Miller[2] : 3 PM = 1·0 : 1·05 : 1·10.

King[17] found RNRZ : ENRZ : MFM : PM = 1·07 : 0·92 : 1 : 0·6.

Stein[18] found RNRZ : Miller[2] : ENRZ = 1·04 : 1·00 : 0·88.

Davidson *et al.*[19] found MFM : '4, 6; 0' code : '6, 8; 0' code = 1·0 : 1·04 : 1·09.

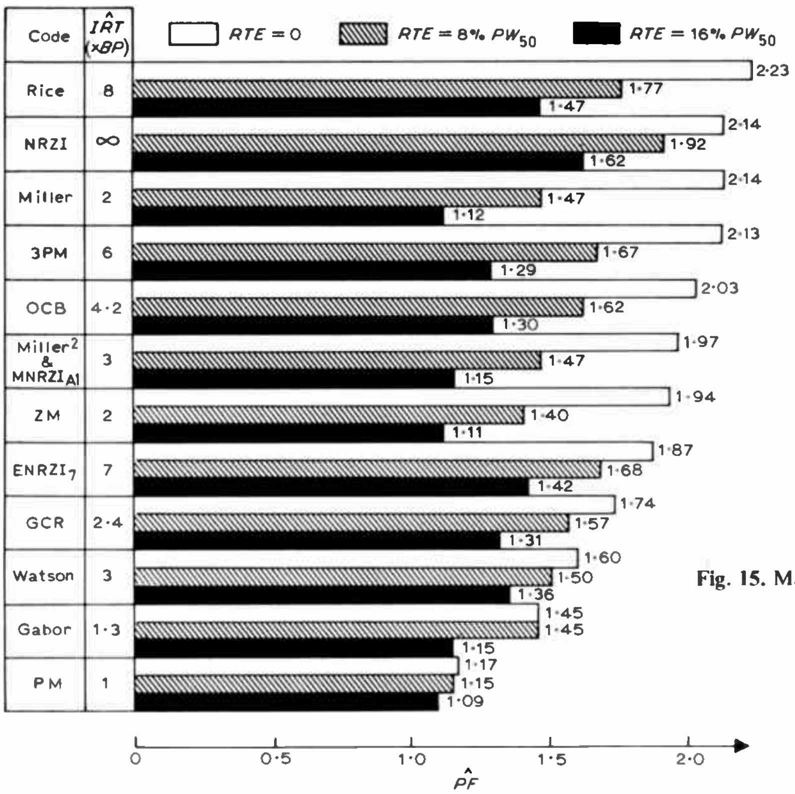


Fig. 15. Magnetic recording codes performance summary.

The clear implication is that, whilst no one agrees which is the best code, there is very little difference between all of the popular codes. Indeed, it has been shown in this study and in the excellent study by King[17] that the choice of detection system can have far more effect on the ultimate performance of a memory than can the choice of the code.

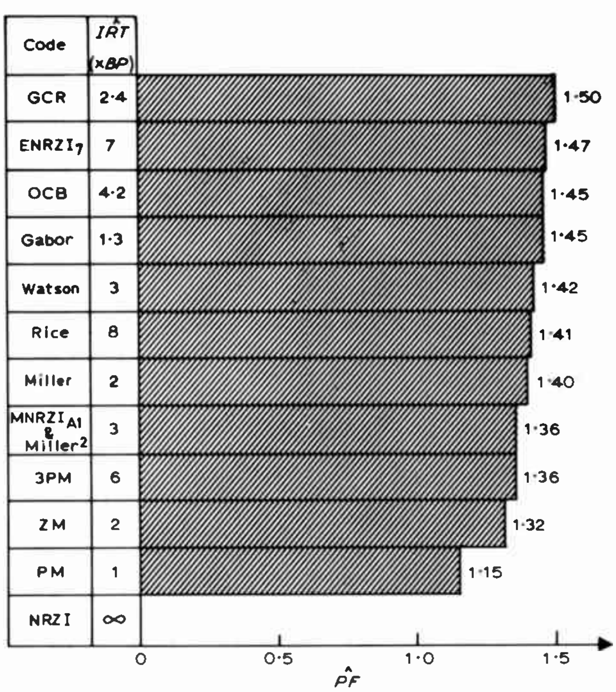
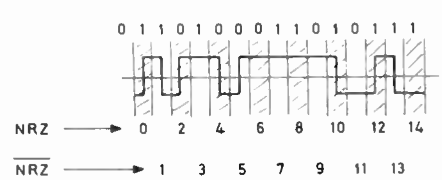
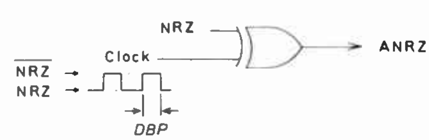


Fig. 16. Performance limits for $RTE \propto \hat{IRT}$.

NRZI remains outstanding for any fixed value of *RTE* and because of this, a new code to emerge recently, Randomized NRZ (RNRZ), has been gaining in acceptance. It attempts to turn NRZ into a self-clocking code by scrambling the data before recording. The idea is that long runs of data without transitions are broken up, and the chances of having a long *IRT* after scrambling are small. It is difficult to see, however, how a typically random data pattern is, in fact, improved by scrambling, i.e. randomizing. If the idea is that data often consist of all 0's, or all 1's (in NRZ), it is very simple to break these up, with much less circuitry than RNRZ requires, by alternating the NRZ definitions, i.e. Alternating NRZ (ANRZ): In even bit periods, code a change from a '0' to a '1' or a '1' to a '0' as a transition at m.b.t.; in odd bit periods, code no change in the data as a transition at m.b.t.:



It can be generated simply by exclusive-ORing the NRZ data with its own clock:



Obviously this code breaks down for an input pattern of 00110011 . . ., but there exists a similar pattern which causes RNRZ to break down.

Although some systems designers are limited in their choice of code by existing standards, e.g. the ECMA and ANSI standards for cassette drives, or the IBM 'standard' for the single density floppy drive, in cases where there is no standard, or where there is one to be set, it behoves the designer to choose a reasonably efficient code. This should not involve much effort, as there are far more important design studies to undertake, such as choice of block structure and detection method, both of which can have more effect on the ultimate formatted capacity of the drive than can haggling over the last few percent achievable by different codes.

Low-cost devices, such as cassette transports and mini diskette drives, often cannot afford the luxury of servo-controlled media speed, or sophisticated encode, decode and detection electronics. For these applications, a code which is very self-clocking, such as PM, should be used. Gabor code could be used for its low $I\hat{R}T$, but the code conversion is quite complex.

To use NRZI in a system really needs a separate dedicated clock track. This is uneconomical in most cases, but for multi-track tape, in either computer data or p.c.m. applications, and certainly for fixed head disks, the overhead involved becomes minimal. The biggest problem is then skew, but l.s.i. chips are now available to overcome this.

For security-conscious applications, the inherent scrambling of Randomized NRZI is an attraction, though the scrambling technique could be applied to any code, at the expense of an extra stage of processing.

In the majority of applications, it is difficult to *really* justify the choice of any particular code, but the original 'compromise' code, Miller code, has much to offer. It is efficient, simple to encode and decode, does not require odd-integral clocks as do some codes, and does not need a sophisticated phase-locked loop.

7 Conclusions

It has been shown how worst-case peak-shift and amplitude patterns can be used to determine the margins in a system. A detailed comparison of many recording codes shows that there is very little to choose between all of the popular ones, as regards the maximum packing density achievable by each.

Pattern-adaptive write (pre-compensation) is seen to be beneficial for some codes (e.g. Watson code), but of little use for most (e.g. NRZI). Similarly, window modification, or pattern-adaptive read, has little to offer. A comparison of detection techniques has shown how they can significantly influence the capabilities of a memory.

The conclusion is that system designers should choose any code which appears suitable for their particular system, and the system should then be designed around the code.

8 References

1 Mackintosh, N. D., 'A superposition-based analysis of pulse-slimming techniques for digital recording', Video and Data Recording Conference, 1979. (IERE Conf. Proc. 43.) *The Radio and Electronic Engineer*, **50**, No. 6, June 1980. (To be published.)
2 Kameyama, T. *et al.*, 'Improvement of recording density by means of cosine equaliser', *IEEE Trans. on Magnetics*, **MAG-12**, no. 6, pp. 746 48, November 1976.
3 Whitehouse, A. E., Ph.D. Thesis, Manchester University, 1970.
4 Dunstan, E. M. and Whitehouse, A. E., 'A 2000 bpi magnetic drum channel with high noise immunity', Conf. on Computer Technology, Manchester 1967. (IEE Conf. Pub. 32.)
5 Tamura, T., *et al.*, 'A coding method in digital magnetic recording', Intermag 1972, *IEEE Trans.*, **MAG-8**, No. 3, pp. 612–14, September 1972.
6 Padalino, M., Workshop on Magnetic Recording, Intermag 1968.
7 Pouliart, W. H. P. and Vandevenne, J. P. H., 'Electrical Intelligence Storage Arrangement', U.S. Patent No. 2,807,004, September 17th, 1957.
8 Woo, W. D., 'Binary Magnetic Recording Apparatus', U.S. Patent No. 3,235,855, February 15th, 1966.
9 Franaszek, P. A., 'Sequence-state methods for run-length-limited coding', *IBM J. Res. & Dev.*, **14**, No. 4, pp. 376–83, July 1970.
10 Gabor, A., 'Adaptive coding for self-clocking recording', *IEEE Trans. on Electronic Computers*, **EC-16**, no. 6, pp. 866–68, 1967.
11 Watson, I., Ph.D. Thesis, Manchester University, 1973.
12 Miller, J. W., U.S. Patent 4,027,335, May 31st 1977.
13 Miller, A., U.S. Patent 3,108,261, October 22nd, 1963.
14 Patel, A. M., 'Zero modulation encoding in magnetic recording', *IBM J. Res. & Dev.*, **19**, no. 4, pp. 366–78, July 1975.
15 Jacoby, G. V., 'A new look-ahead code for increased data density', *IEEE Trans.*, **MAG-13**, no. 5, pp. 1202 4, September 1977.
16 Huber, W. D., 'Maximization of lineal recording density', *IEEE Trans.*, **MAG-13**, no. 5, pp. 1208–10, September, 1977.
17 King, D. A., 'Comparison of pcm codes for direct recording', Intern. Telemetering Conf., September 1976, pp. 526–40.
18 Stein, J. H., 'High Density Recording: Facts relating to Standardization', BSI Doc. no. 78/60564.
19 Davidson, M., *et al.*, 'High density magnetic recording using digital block codes of low disparity', *IEEE Trans.*, **MAG-12**, no. 5, pp. 584–86, September 1976.

Manuscript received by the Institution on 2nd April 1979
(Paper No. 1928/Comp 195)

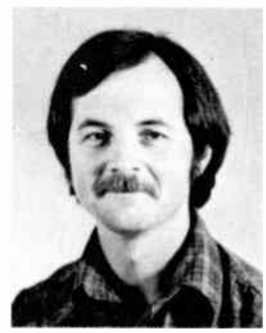
Contributors to this issue

Ralph Benjamin (Fellow 1975) has been Chief Scientist and Superintending Director at the Government Communications Headquarters since 1971. A graduate of the University of London, he joined the Royal Naval Scientific Service in 1944 and after individual merit promotions to Senior Principal Scientific Officer and Deputy Chief Scientific Officer, he was appointed Head of Research and Deputy Director of the Admiralty Surface Weapons Establishment in 1961. In 1964 he transferred to the Admiralty Underwater Weapons Establishment as Director and Chief Scientist, also becoming a qualified Naval Diving Officer. From 1965 to 1971 he combined these appointments with that of Director of Underwater Weapons R & D at the Ministry of Defence. In 1964 London University awarded Dr Benjamin the Ph.D. degree for a thesis on signal processing, and in 1970 the D.Sc. for his contributions to general electronics. He has published a book on 'Modulation, Resolution and Signal Processing in Radar, Sonar and Related Systems' and is author of numerous papers in this Journal and elsewhere.

He was appointed C.B. in the recent New Year Honours List.

Neil MacKinnon (Fellow 1979, Member 1970) studied at Worcester College, Oxford, and the Royal Military College of Science whilst serving in the Royal Corps of Signals. During his military service he spent eight years in electronic warfare environments and other appointments included those of Airborne Forces Development Officer for electronic devices such as transponders and locator systems, and Weapons Staff Officer at SRDE, Christchurch working on the Clansman project. After leaving the Army in 1973 Mr MacKinnon spent three years with the Racal Electronics Group during the formative period of their communication security company, Racal Datacom, to whom he is now a Consultant. His current interests are in the application of analogue solid-state devices, such as c.c.d.s, to signal processing, and the system planning of secure diplomatic radio networks.

'Nick' Kouvaras graduated first from the Athens Higher School of Electronics Engineering, after which he attended the Faculty of Mathematics of the University of Athens and obtained his B.Sc. degree in mathematics. In 1960 he joined the Electronics Department of the 'Demokritos' Nuclear Research Centre, and he has done research and development work on digital systems which has led to several publications including a paper in this Journal in September 1978 on delta modulation; he is also involved in a programme of work on digital processing systems for speech.



Nigel Mackintosh received his B.Sc. in computer science at the University of Manchester in 1971 and his subsequent work on magnetic aspects of the MU5 research computer gained him his Ph.D. in 1975. At Burroughs Machines in Scotland he helped develop the world's first 1 Mbyte floppy-disk drive, and later at Racal Recorders in Southampton he applied microprocessor technology to communications recorders and equipment. Dr Mackintosh is now with the Advanced Techniques Group of Burroughs Machines in California evaluating the limits of digital data recording.

Ian Vance joined Standard Telecommunication Laboratories in 1968 after receiving the degree of Bachelor of Engineering from the University of Liverpool. In 1973 he was admitted as Master of Science at Aston University. Mr Vance has worked in many aspects of the field of radio communication and microwaves, leading in recent years to an interest in highly integrated forms of such equipment. He is currently manager of the Radio Frequency Sub-Systems Department at the Harlow establishment of STL.