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 51 No. 11/12 November/December 1981

The Radio and Electronic Engineer

The Journal of the Institution of Electronic and Radio Engineers

On Becoming Old-Fashioned

It has been remarked on more than one occasion that any electronics engineer who does not disagree profoundly with the opinions of engineers from other disciplines is becoming old-fashioned or senile and should be pensioned off. Some of those most directly concerned with the shaping of IERE policy have therefore been somewhat alarmed to discover that they were unable to quarrel violently with the recent statements by the Department of Industry and the Department of Education and Science about the Government's post-Finniston intentions. In respect of the first, of course, the cautionary 'there's many a slip' is still valid. The proposed British Engineering Council cannot undertake the functions proposed for it unless CEI is prepared to surrender its Charter: this means that it must convince a majority of the 198,000 Chartered Engineers who would be required to approve that surrender that it will be able to offer more than good intentions.

The DES proposals (which were summarized in *The Electronics Engineer* of 3rd September 1981), on the other hand, can be implemented without prior approval by any other body. We were therefore pleased to note that they conform very closely to the recommendations made by this Institution in its written submissions to both the Finniston Inquiry and the subsequent National Conference on Engineering Education and Training. It is especially encouraging to note that the Department of Education appears to have heeded the many protests about the Report on Conference Theme 6, namely Technician Education, and now states that views differed considerably on the subject of the proposed practically-oriented BTech for Technician Engineers. As some readers will know, the Trent Conference of Heads of Polytechnic Engineering Departments latched on to the idea and proposed a BSc(Eng) of the kind envisaged and attempts have been made to 'sell' it as a replacement for the HND in Electrical and Electronic Engineering. We were pleased to learn that the Joint Committee concerned decided it would not be a satisfactory replacement: less so to be told that the DES is so anxious to wind up the Joint Committees that an award of the HND kind can only continue in being beyond the present decade if the Technician Education Council can be persuaded to formulate and offer one.

To some extent, therefore, the fact that this editorial is not a vehement protest is indicative of our recognition that protest can no longer achieve anything. Too much that we saw as of value in the Joint Committee pattern of awards has already been sacrificed in the name of progress or uniformity: the little that remains now seems doomed to follow—possibly because the die was cast at a time when engineers in other branches of the profession could not see what those electronics boffins were so worried about.

Increasingly however, in discussion with those engineers, we hear of growing concern about the uncertainties which have stemmed from the changes in the pattern of education instigated by misguided equalitarians who cannot understand that sorting must take place at some stage in the individual's education, and that to delay it too long is at best wasteful and at worst harmful. So perhaps the idea that an electronics engineer who was not something of a maverick was 'past it' is itself outdated. Possibly, the increasing use of electronic devices in all fields of engineering has led to a greater awareness among other engineers of the vast potential which electronics holds for 'engineering our future' and the better understanding means not that we are joining them, but that they are joining us.

K.J.C.

Members' Appointments

CORPORATE MEMBERS

T. O. Achebe (Member 1965, Graduate 1957) who has been Chief Engineer with Nigerian Television since 1978, has taken up an appointment as Director of Engineering, Anambra Television, Enugu.

G. D. Bishop, B.A., M.Sc. (Member 1977), Head of the Department of Engineering at Moston College of Further Education, Manchester since 1977, has moved to Bridgwater College in Somerset where he has taken up the post of Vice Principal.

Lt **J. Bruce, B.Sc., RN** (Member 1981, Graduate 1975) has been appointed to the Royal Navy Polaris School prior to taking up an appointment as the Weapon Engineer Officer of a Polaris Submarine.

S. S. Duncan (Member 1957) has formed Spencer Duncan & Associates to provide consultancy services in noise and vibration control engineering and instrumentation. He moved to Worcestershire after 21 years at Hawker Siddeley Dynamics where as Chief Environmental Engineer, he was responsible for noise and vibration instrumentation and engineering for several major missile projects.

K. D. Fishenden, M.Sc. (Member 1974, Graduate 1969) has been promoted to Engineering Services Manager of De La Rue Systems, Portsmouth. Mr Fishenden previously held appointments as Senior Project Manager and Programme Manager with Crosfield Electronics, also part of the De La Rue Group.

M. J. Furniss (Member 1973, Graduate 1967) has been promoted to Senior Principal Scientific Officer in the Microelectronics Unit, Research and Development Division at the British Railways Board, Railway Technical Centre, Derby.

N. F. W. King (Member 1968, Graduate 1963) has been appointed Managing Director of RS Components. Previously Marketing Director, he was appointed a Main Board Director of Electrocomponents, the parent company, in 1978.

R. M. G. Maule, B.Sc. (Member 1971, Graduate 1968) is now Area Sales Manager, North Asia/Pacific with Motorola Communications International in Hong Kong. He was previously with the company in Japan.

D. W. F. Medcraft (Member 1963), formerly head of the Digital Data Networks Division, Telecomms Headquarters, British Post Office, has been appointed Deputy Director Overseas Liaison and Consultancy Department and General Manager, British Telconsult.

S. E. Osime (Member 1974, Graduate 1971) has been promoted to Assistant Chief Engineer with the Nigerian Posts and Telecommunications Department in Benin City.

B. S. Plumb, B.Sc., M.Sc., Dip. E.E. (Member 1979, Graduate 1969) has been promoted to Principal Lecturer and Director of Studies for the B.Sc. degree course in electronic systems and control engineering at Sheffield Polytechnic.

S. A. Seriki (Member 1974, Graduate 1970) is now Head of Engineering Studies at the P & T Training Centre in Oshodi, Nigeria. He was previously Principal Engineer and Head of the Radio School.

M. J. Whiteman (Member 1971) is now Senior Instrument Co-ordinator with Pratt & Whitney Aircraft of Canada, at Mississauga, Ontario. Before going to Canada in May 1981, he was a Professional & Technology Officer II in the Ministry of Defence.

NON-CORPORATE MEMBERS

G. D. Brown (Graduate 1970) who has been with Geophysical Service International since 1971, with the position of South East Asia Marine Data Collection Manager since February 1981, has now been appointed Manager, Marine Operations for Australia and New Zealand, based in Perth.

H. G. F. S. Parish, M.Sc. (Graduate 1968) is now Chief Scientist and Head of the Systems Engineering Sub-division within the Evaluation Division of the Licensing Branch of the Atomic Energy Board in Pretoria.

Obituary

The Institution has learned only recently of the deaths of the following members.

Roy Walker, B. Eng. (Member 1967) of Toowoomba, Queensland, died in July 1980, aged 52. After graduating at the University of Sheffield, Mr Walker carried out research on losses in magnetic materials in its Department of Electrical Engineering for three years until 1953 when he served two years National Service with the RAF as a Ground Radar Fitter. From 1955 to 1967 he was with de Havilland Propellers, Hatfield (subsequently Hawker Siddeley Dynamics) initially as a Senior Electronics Engineer and finally as an Assistant Group Leader. In 1968 Mr Walker was appointed Senior Lecturer in electronics and control at Sheffield Polytechnic and two years later went to Darling Downs Institute of Advanced Education, Toowoomba, Queensland, as a Senior Lecturer, a position he held at the time of his death.

Peter Jordan Watson (Member 1971) of Rochdale, Lancashire, died in November 1980, aged 40. Following an electrical apprenticeship with the National Coal Board in Co. Durham, Mr Watson first worked for Reyrolle and Company, Hebburn, as a tester of electrical equipment. He then joined Burndept and Co at South Shields in 1962 as an electronics engineer and two years later moved to Applied Research and Engineering at Peterlee as a Development Engineer. From 1960 until 1976 he was with GEC Measurements at Stafford as a Project Leader and in 1977 he was appointed Chief Engineer of the Instrument Department in the Instrumentation Division of Ferranti at Moston.

Benjamin James Veneear (Member 1958, Graduate 1951, Student 1943) of Kirkcaldy, Fife, died on 10th April 1981, aged 57. After leaving school Mr Veneear worked for two years with Standard Telephones and Cables, Enfield, and during the war served with the Royal Navy. On demobilization he became a leading draughtsman with Edison Swan and while with this company, he studied at Northampton Engineering College gaining a B.Sc. degree of London University in electrical engineering. From 1951–52 he was Laboratory Manager with British Physical Laboratories, Radlett and he subsequently held positions of Works Manager at Fortiphone, Camberwell, and Lines Bros, Richmond, Surrey, before joining Tape Recorders, Tottenham, as Director responsible for engineering design and production. In 1965, Mr Veneear was appointed Technical Director of Highland Electronics at Dunfermline, Fife, with whom he remained until his sudden death.

John Henry Volker (Associate Member 1973, Associate 1968) of Merstham, Surrey, died last year, aged 69. Born and educated in the United States, Mr Volker joined the US Navy in 1930 as an electronics technician. He was commissioned after the outbreak of war and continued in the service until 1957, his last appointment being that of Liaison Officer, Electronics, at the US Office of Naval Research in London with the rank of Lieutenant Commander. From 1958–64, he was in business on his own account at Hookwood, Surrey, being concerned with electrical, mechanical and agricultural engineering products. For the next three years he was Plant Engineering Manager for a ready-mixed concrete company, devising and supervising control and weighing equipment In 1967 he joined Foxboro-Yoxall of Redhill as lecturer on electronic and pneumatic control instruments and he remained with the company until his retirement in 1977.

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.

A Professional Union: The Evolution of the Institution of Professional Civil Servants
JAMES E. MORTIMER and VALERIE A. ELLIS. George Allen & Unwin, London, 1980. 16 × 24.5 cm. 450 pages. £15 (hardback).
CONTENTS: The birth of the Institution and the social background. Thwarted expectations, 1919–25. The development of the Institution, 1919–25. The general strike and the civil service unions. From the general strike to the Tomlin Commission. The Tomlin Commission and the Carpenter Committee. The years of depression. The second half of the 1930s. The early years of the second world war and a new leadership for the Institution. The war effort and a new social role and structure for the Institution, 1942–5. Postwar reorganization. From the 1940s to the 1950s: Some basic problems remain. The 'purge' and civil rights.

This specially commissioned history commemorates the first 60 years of the Institution of Professional Civil Servants, which has over 100 000 members throughout the professional and specialist grades of the UK Civil Service. The IPCS is the largest single body of engineers and research scientists in the country and has played a central part in the debates about the role of scientific and technical developments, nuclear energy, etc.

Electronics for the Service Engineer. Part 2
IAN F. SINCLAIR (*Braintree College of Further Education*). Technical Press, Oxford, 1980. 18.5 × 24 cm. 271 pages. £4.95.
CONTENTS: Basic principles and circuitry. Domestic electronic equipment. Industrial and telecommunications circuits.

These volumes are intended to serve as an introduction to electronic circuit operation and fault-finding procedure for the service technician and the interested amateur. In addition to its primary purpose as a textbook for the City & Guilds of London Institute course the present volume covers material which would be valuable for students following a Technician Education Council course at level II.

Test Gear Projects
TERRY DIXON. Macmillan, London, 1980. 13.5 × 21.5 cm. 101 pages. £3.95 (paperback).
CONTENTS: Power supplies. Generators and signal injectors. Test meters. Component checkers. Oscilloscope accessories. Workshop gadgets.

There are more than 30 projects outlined including a selection of power supplies, signal injectors, a reference oscillator, noise generators, a logic probe, multimeter, capacitance bridge, transistor tester, oscilloscope calibrator, and an oscilloscope dual trace adapter. Full constructional details are given.

Frequency Synthesizers: Theory and Design (2nd Edition)
VADIM MANASSEWITSCH (*Engineering Consultant*). Wiley, New York, 1980. 16 × 23 cm. 582 pages. £19.25.
CONTENTS: Frequency synthesis. System analysis. Shielding. Analog phase-locked loops. Digital phase-locked loops. Basic circuits. Frequency synthesizers. Frequency reference sources. Troubleshooting of synthesizers.

Electronic Testing and Fault Diagnosis
G. C. LOVEDAY (*Havering Technical College*). Pitman, London, 1980. 18.5 × 24 cm. 212 pages. £5.00.
CONTENTS: Specifications. Reliability. Electronic components. Digital logic circuits. Circuits using linear integrated circuits. Power supply and power control circuits. System maintenance and fault diagnosis.

An introduction to electronic testing and fault diagnosis for higher technician students using essentially a non-mathematical approach. One of its aims is to combine the subjects of testing, reliability and fault diagnosis.

Handbook of Filter Design
RUDOLPH SAAL (*Technical University of Munich*). AEG-Telefunken, Ulm, 1979. 17.5 × 24.5 cm. 800 pages. 220 DM.
CONTENTS: General remarks on filter design. LC lowpass filter. LC filter circuits by transformation of a lowpass filter. Frequency band-separation circuit. Frequency transformations for filter circuits. RC active filters. Digital filters. Numerical examples.

Contains design data and tables essential to solve a broad range of filter design problems. The application of the tables is explained in detail and more than 25 worked examples are given for illustration. Text is in both German and English.

Electrical & Electronic Principles
NOEL MORRIS (*North Staffordshire Polytechnic*). Pitman, London, 1980. 18.5 × 24.5 cm. 230 pages. £4.95.
CONTENTS: Fundamentals of electricity. Ohm's law, power and energy. Series and parallel circuits containing resistors. Kirchhoff's laws. The magnetic field. Electromagnetism. Electrostatics. Direct current switching circuits. Alternating voltage and current. The single-phase transformer. Measuring instruments.

Intended mainly for technicians but not angled towards any specific course or examination.

Antennas in Matter
RONOLD W. P. KING (*Howard University*) and GLENN S. SMITH (*Georgia Institute of Technology*) MIT Press, Cambridge, Mass., 1981. 16 × 23.5 cm. 868 pages. £46.50
CONTENTS: Fundamentals. The insulated antenna in its various forms. The bare antenna in air and dissipative media. The linear antenna as a probe in material media. Waves and antennas in a dissipative half-space. Advanced theory. Electromagnetic theory and constitutive relations. Constitutive relations and parameters. The bare linear antenna in a homogeneous isotropic medium. Theory of the insulated antenna. The circular-loop antenna in a material medium. The circular-loop antenna with a spherical insulation. Antennas near a planar interface Experimental models and measuring techniques. Construction of experimental models. Measuring techniques and apparatus.

This book provides an introduction to antennas in both dissipative and dielectric media on the level of the undergraduate electrical engineer or physicist without sacrificing accuracy or rigour. Part I is introductory, elementary and fundamental. Part II is an advanced treatment of antennas in various media. Part III is concerned with the construction of experimental models and the techniques of measurement relating to antennas and probes in a general dissipative or dielectric medium.

Electronic Devices and Components
J. SEYMOUR (*Thames Polytechnic*) Pitman, London, 1981 15 × 22.5 cm. 504 pages. £9.95
CONTENTS: Electrons in atoms. Electrons in crystals. Contacts between materials and p-n junctions. Bipolar junction transistors and thyristors. Optoelectronic devices. Field-effect transistors and charge transfer devices. Integrated circuits. Vacuum and gas-filled devices. Microwave devices and electrical noise. Dielectric materials and components. Magnetic materials and components.

This book provides an introduction to the physical principles underlying the operation of present-day electronic devices and components. It is written to provide a complete course which can extend from the first to the final year of a degree in electrical and electronic engineering or applied physics.

Electric Machines and Transformers
LEONARD R. ANDERSON (*Southern Alberta Institute of Technology*) Prentice-Hall, Englewood Cliffs, 1981. 18 × 24 cm. 305 pages. $18.95
CONTENTS: Direct-current generators. Direct current motors. Transformers. Polyphase induction motors. Polyphase synchronous alternators. Polyphase synchronous motors. Single phase motors. The mechanics of electric motor drives. General information on motors and electric power systems. Motor starters and controllers.

Letters to the Editor

From: **J. K. Murray,** T.Eng. (CEI) (Associate Member)
G. M. Voglis, Ph.D.;
R. B. Mitson, M.Sc., C.Eng., M.I.E.R.E., and
A. R. Pratt, Ph.D.

James Clerk Maxwell

I have just received the September 1981 issue of the IERE Journal and read with interest your Editorial entitled, 'James Clerk Maxwell 1831–1879". My reason for writing is for the benefit of readers who might wish to visit the grave of James Clerk Maxwell. His grave is to be found in the Old Kirk at Parton in Kirkcudbrightshire, not at Glenlair as stated in the Editorial. Several of my ancestors were buried alongside his grave, so I am sure of my facts.

J. K. MURRAY

51 Colney Heath Lane,
St Albans, Hertfordshire
16th September 1981

Editorial Note

We are grateful to Mr Murray for drawing attention to this geographical error. Fuller details of James Clerk Maxwell's resting place and memorials are given in Professor R. V. Jones' Address published in this issue.

Developments in Sector Scanning Sonar

I am writing in connection with a publication by Holley, Mitson and Pratt entitled 'Developments in sector scanning sonar'[1] which has recently come to my notice.

The paper represents an attempt to duplicate the scanning sonar known as the A.R.L. Scanner, replacing the valve technology originally used with its contemporary solid-state counterpart. As a background, it is relevant to mention that this equipment is based on 'Modulation Scanning', a unique technique first introduced over 30 years ago at the Admiralty Research Laboratory, Teddington (now renamed the Admiralty Marine Technology Establishment) and patented in 1956[2]. Progressive research and development combined with experience gained at sea led, in the late '50s, to the design and contruction of an advanced scanning sonar named at the time the 'Bifocal Equipment' and later renamed as indicated earlier, the A.R.L. Scanner. Extensive evaluation at sea in the early '60s soon established this scanning sonar as the leader in the field[3], a position maintained to date. In fact, to my knowledge, all sonar scanning equipments currently in use in this country and on the Continent and Europe, both military and non-military, are based exclusively on the 'Modulation Scanning' techniques referred to above, all other competitive methods having either fallen into oblivion[4,5] or not applied[6] because of their serious inherent fundamental limitations and other disadvantages.

Following partial declassification of this work, coinciding with my retirement from the Service, I published in the early '70s a series of papers dealing with the theoretical basis of 'Modulating Scanning'[7] and also design details of the A.R.L. Scanner[8] representing a distillation of all development work and operational experience. Inevitably the paper referred to in the outset[1] leans heavily on my work, yet the authors deemed it appropriate to criticize the patent by claiming in page 141, 3rd paragraph, 3rd sentence that, quote, 'an incomplete list of methods is given in Voglis *et al.*, i.e. the patent, without referring explicitly to method(s) omitted. Instead the authors follow this statement with a sentence unambiguously implying that the arrangements (A), (B) and (C) of Fig. 1, page 142, represent 'other methods' not covered by the patent.

I wish to put on record my objections to this unjustified claim and to assert (i) that no method based on 'Modulation Scanning' other than those described in the patent or my published work, has so far been suggested or applied and (ii) that, of the three diagrams (A), (B) and (C) of Fig. 1, page 142, (A) is an unsuitable application and (B) and (C) are incorrect copies of the methods put forward in the patent or my published work.

I shall be grateful for the publication of this letter aimed at setting the record straight.

G. M. VOGLIS

'Lakeside', Onslow Road,
Burwood Park, Walton-on-Thames, Surrey,
26th May 1981

We are indebted to Dr. Voglis for his comments and for this opportunity to expand upon the comment that we made in the 1975 paper.[1]

The primary context of this was that the 1956 patent contained an incomplete list of suitable modulation methods. This view was in the main due to the publication of Voelcker's work[9] and his extensive list of references to earlier work. These papers contain a series of modulation methods not used by Voglis which may be worth contemplating for sector scanning applications. The second context of our statement was that by 1975 there were a number of other modulation methods which could be used to mechanize sector scanning requirements. We have taken the class of modulation scanning to include all those methods which use modulation as an essential element in satisfying the sector scanning signal processing. Techniques involving modulation in a secondary role (e.g. for frequency changing) are therefore excluded from the class. References 4 and 6 are by this token modulation scanning methods as is also reference 10. We feel that a detailed explanation of these techniques is outside the scope of a letter to the Editor. Interested readers may verify for themselves that all of these techniques use modulation as an essential element in the signal processor and are thus modulation scanning techniques.

It is not relevant to us whether these methods have either fallen into oblivion or have not been applied. In either case, they only await some new twist of technology to make then the preferred method. For example, the work of Copping in reference 6 makes use of a direct Fourier Transform approach. The limitations on the performance versus cost made this method unattractive in the mid 1970s. However, cost predictions of $10 for a single integrated circuit Fourier Transform processor with acceptable performance[11] may make this the preferred method of the late 1980s. None of the methods we have mentioned have any inherent fundamental limitations which cause them to be unfit for use as modulation scanning processors.

Dr. Voglis is correct in pointing out that the arrangements (A), (B) and (C) are indeed taken from his work. Unfortunately, an uncorrected typographical error changed 'some of these methods' into 'some other methods'. We regret any misunderstanding or misinterpretation which has been caused by this or any other typographical errors appearing in our paper.

A. R. PRATT

Polytechnic Marine,
Royal Oak Way, Daventry,
Warwickshire.
2nd September 1981

R. B. MITSON

Fisheries Laboratory,
M.A.F.F.,
Lowestoft, Suffolk NR33 0HT

References

1 Holley, M. L., Mitson, R. B. and Pratt, A. R., 'Developments in sector scanning sonar', Conference on Instrumentation in Oceanography, IERE Conference Proceedings No. 32, pp. 139–53 1975.
2 Voglis, G. M., Willis, H. F. and Buckingham, J., British Patent No. 926580, Application date 10th April 1956.
3 Voglis, G. M. and Cook, J. C., 'Underwater applications of an advanced acoustic scanning equipment', *Ultrasonics*, **4**, pp. 1–9, January 1966.
4 Tucker, D. G., Welsby, V. G., and Kendall, R., 'Electronic sector scanning.' *J. Brit. Instn Engrs*, **18**, pp. 465–84, August 1958.
5 Nairn, D., 'Theoretical possibilities of digital sonar systems', Proceedings of Symposium on Signal Processing in Radar and Sonar Directional Systems, Paper No. 20, IERE, July 1964.
6 Copping, A. J., 'Electronic sector scanning. Multiple beam-forming networks', Ph.D. Thesis, Loughborough University of Technology, 1973.
7 Voglis, G. M., 'A general treatment of modulation scanning as applied to acoustic linear arrays', (Part 1), *Ultrasonics*, **9**, pp. 142–53, July 1971; (Part 2), **9**, pp. 215–23, October 1971.
8 Voglis, G. M., 'Design features of advanced scanning sonars based on modulation scanning', (Part 1), *Ultrasonics*, **10**, pp. 16–25, January 1972; (Part 2), **10**, pp. 103–13, May 1972.
9 Voelcker, H. B., 'Toward a unified theory of modulation', *Proc. IEEE*, (Part I), **54**, no. 5, pp. 735–55, May 1966.
10 Squire, W. D. and Whitehouse, H. J., 'Convolutionally scanned multi-dimensional arrays', Proc. of the Satellite Symposium on Underwater Acoustics, pp. 2.3/12, August 1974, Birmingham.
11 Ashman, R. H., 'Digitisation is on the way for FFT designs', *Electronic Engineers Design Magazine*, p. 99, August 5th, 1981.

Standard Frequency and Time Service

Communication from the National Physical Laboratory

Relative Phase Readings in Microseconds NPL—Station

(Readings at 1500 UTC)

AUGUST 1981	MSF 60 kHz	GBR 16 kHz	Droitwich 200 kHz
1	−0·1	15·0	72·5
2	−0·3	14·7	—
3	−0·3	14·7	—
4	−0·5	14·6	72·5
5	−0·7	14·6	72·5
6	−1·4	14·7	72.5
7	−1·2	14·6	72.6
8	−1·2	14·5	72·6
9	−1·0	14·6	—
10	−1·2	14·8	72.8
11	−1·4	15·1	73·0
12	−1·4	14·7	73·0
13	−1·4	14·8	73·1
14	−1·6	14·3	73·2
15	−1·6	14·6	73·3
16	−1·4	14.5	73·4
17	−1·2	14·6	73·4
18	−1·4	14·5	73·6
19	−1·4	14·5	73·6
20	−1·4	14·1	73·7
21	−1·4	14·6	73.8
22	−1·4	14·5	73·8
23	−1·4	14·1	73·8
24	−1·4	14.2	73.9
25	−1.6	14·3	73·9
26	−1·8	14·2	73·9
27	−1·8	14·3	73·9
28	−1·8	14·1	74·0
29	−1·8	13·3	73·9
30	−1·8	14·2	73·9
31	−1·6	13·3	73·9

SEPTEMBER 1981	MSF 60 kHz	GBR 16 kHz	Droitwich 200 kHz
1	−1·6	13·0	73·9
2	−1·7	13·8	73·8
3	−1·8	13·9	73·7
4	−1·7	14.0	73·7
5	−2·0	14·5	73·7
6	−2·0	14·3	73·7
7	−2·0	13·9	73·7
8	−2·2	13·5	73·8
9	−2·2	13·8	73·8
10	−2·4	13·3	73·8
11	−2·4	13·8	73·8
12	−2·4	13·8	73·9
13	−2·5	13·3	74·0
14	−2·4	13·3	74·1
15	−2·6	13·0	74·3
16	−2·6	13·3	74·4
17	−2·8	13·1	74·5
18	−2·9	13·3	74·4
19	−2·6	—	74·5
20	−2·6	—	74·4
21	−2·7	—	74·5
22	−2·7	—	74·6
23	−2·9	—	74·7
24	−3·0	—	74·8
25	−2·9	—	74·8
26	−3·1	13·1	74·9
27	−3·0	13·1	75·0
28	−3·1	13·2	75·1
29	−3·1	13·7	75·1
30	−3·3	13·6	75·3

Notes: (a) Relative to UTC scale (UTC_{NPL}–Station) = +10 at 1500 UT, 1st January 1977.
(b) The convention followed is that a decrease in phase reading represents an increase in frequency.
(c) 1 μs represents a frequency change of 1 part in 10^{11} per day.
(d) It may be assumed that the satellite stations on 200 kHz at Westerglen and Burghead will follow the day to day changes in these phase values.

Correction: July 1981.

The MSF readings for July, published in the October Journal, were in error by a constant +6 microseconds. To obtain the correct values *subtract* this figure from each reading given, which will thus range from +0.9 μs to −0.1 μs. All the other readings i.e. those for GBR and Droitwich are unchanged.

New and Revised British Standards

Copies of British Standards may be obtained from BSI Sales Department, 101 Pentonville Road, London N1 9ND. Non-members should send remittances with orders. Subscribing members will be invoiced and receive 50% discount.

ELECTRONIC TUBES AND VALVES

The recent decision by CENELEC (European Committee for Electrotechnical Standardization) to harmonize British and European Standard dimensions for electronic tubes and valves has given BSI an opportunity to revise the relevant British Standard, BS 448 **Dimensions of electronic tubes and valves** which is to be published in two parts.

The first part of the revision is now available as Part 1 **IEC dimensions** (£25.00) this being identical with IEC 67 and its four supplements. It provides a range of base and outline drawings, and conversion data for internationally-used electronic tubes and valves. Part 2, which will follow in a few months' time, will cover bases and gauges that are a British national requirement and are not contained in Part 1.

ELECTRONICS LETTER SYMBOLS

A British Standard of commercial and technical significance to the electronics component and equipment manufacturing industries has been updated by BSI and aligned with IEC 148. This is BS 3363 **Letter symbols for semiconductor devices and integrated microcircuits,** (£16.50) which supersedes the 1968 edition (not withdrawn).

The availability of an internationally agreed system of letter symbols is, of course, a great advantage to those involved in mathematical calculations covering the operation of semiconductors and integrated microcircuits. The main purpose of the standard, however, is to enable the technical characteristics of manufactured products to be clearly described to potential customers, thus enhancing trading prospects.

Supplement 1 (£12.00) provides new letter symbols for bipolar transistors, field effect transistors, voltage reference and voltage regulator diodes, digital integrated microcircuits, analogue integrated circuits and mixer diodes. (Equivalent to IEC 148A).

Supplement 2 (£9.00) gives new letter symbols for bipolar transistors, low-power signal diodes, voltage reference and voltage regulator diodes, rectifier diodes, thyristors, field effect transistors, digital and analogue integrated microcircuits, mixer diodes and current regulator diodes. (Equivalent to IEC 148B). There is unlikely to be another supplement for some two to three years.

ANALOGUE COMPUTING

A further standard in the twenty-part series, BS 3527 *Glossary of terms used in data processing*, is entitled Part 19 **Analog computing** (£5.00). Identical with ISO 2382, Section 19, it defines current data processing concepts in relation to analogue and hybrid arithmetic units, function generators, converters and modes of operation for such components. The glossary presents selected terms and definitions which are drafted to avoid as far as possible any peculiarities associated with a particular language. All the main areas of data processing are dealt with including the principal processes and types of equipment used, the representation, organization and presentation of data, the programming and operation of computers, input-output devices and peripheral equipment as well as individual applications.

MACHINE TOOL SYMBOLS

BSI has published a revision of BS 3641 **Symbols for machine tools** Part 2 **Numerical Control** (£10.40), which is generally in accordance with ISO 2972 issued by the International Organization for Standardization, but retains a number of national symbols that appeared in the original 1972 edition. Where difference exist between the two documents they are clearly indicated in the foreword.

The revised standard specifies basic designs for symbols intended for use on indicator plates of numerically-controlled machine tools. Drawings relate only to the form of symbols and are not absolute dimensional representations. The standard gives recommendations for proportions to be used in the design of symbols, independent of actual dimensions, and provides for the grouping of symbols.

UNIFORM FRAME AND MOUNTING DIMENSIONS FOR LOUDSPEAKERS

Another part of the British Standard for specifying and measuring the characteristics of sound system equipment has just been published. Part 5 Section 5.1 of BS 5428 **Dimensional characteristics of single moving-coil (dynamic) loudspeakers,** (£7.50) like the other parts of this standard, is of specific interest to manufacturers of electronic and audio equipment, and will be a useful guide to them when introducing new types of loudspeakers.

The object of the standard is to secure as great a measure of interchangeability as seems practicable, and to discourage unnecessary divergencies, while at the same time allowing the maximum liberty of choice in the design of the loudspeakers, cabinets and enclosures. The standard is limited to the outer dimensions of the frame and the mounting dimensions, for example, the size and disposition of fixing holes, and it excludes the cone dimensions and the depth of the loudspeakers. BS 5428 is identical with IEC 268–14.

ELECTRICITY AND MAGNETISM

BS 5775 **Quantities, units and symbols** Part 5 **Electricity and magnetism** (£12.00) completes a further stage of the BSI programme to reproduce all parts of the equivalent international standard (ISO 31) published by the International Organization for Standardization.

The first six parts of the British version were published in 1979 and deal with space and time, periodic phenomena, mechanics, heat, acoustics and mathematical symbols. Each part includes a tabulation of the more important physical quantities in the appropriate discipline, giving brief definitions, letter symbols and the relevant SI units of measurement. Subsequent parts will be published as the ISO work progresses, though not necessarily in numerical order.

AUDIO-VISUAL, VIDEO AND TELEVISION SYSTEMS

Four parts of a new British Standard of particular interest to manufacturers and users of visual aids and television for educational, training and similar applications, have recently been published. BS 5817 **Audio-visual, video and television equipment systems** is being planned in twelve parts, and is designed as a framework into which further items can be fitted with the advent of new technologies in this field. The systems covered in the standard range from very simple general purpose devices to highly professional equipment used, for example, for audio-visual distribution systems.

The standard aims to make it easier to determine the quality of audio-visual apparatus and to compare different types of apparatus and decide on their proper application, by listing the characteristics which are useful for their specification.

BS 5817 Part 1 **General** (£7.50); Part 2 **Definitions** (£5.00), and Part 10 **Audio cassette systems** (£7.50) appeared earlier this year, while the latest document to be published is Part 8 **Symbols and identification**, (£7.50) which specifies symbols relating to controls and input, output, and remote control connectors and terminals for use in education and training. They are selected from IEC Publication 417. A supplement now in preparation will cover a number of additional functions for which no symbols yet exist. A future section of Part 8 will be devoted to the marking of equipment rating plates with details of appropriate supply voltages.

Other parts of the standard which are in the pipeline will deal with such things as connectors; electrical ratings; control, synchronization and address codes; safety requirements (a guide to the safe handling and operation of audio-visual equipment with particular reference to system safety and cross-references to other publications on equipment safety); methods for measuring and reporting the performance of audio-visual equipment, etc.

BS 5817 is being developed on identical lines to the international standard on this subject—IEC 574.

FIRST COMPUTER LANGUAGE STANDARDS

The first British Standards to deal with general-purpose programming languages, have been published and specify the semantics and syntax for two different computer programming languages (RTL/2 and CORAL 66) already in common use for the control of industrial processing and general applications in real-time computer systems.

The first standard is BS 5904 **Computer programming language RTL/2** (£14.50), which covers a method for writing both application and system programs for use in real-time computing and is especially suitable for on-line data acquisition, communication and control systems. The overall objectives of the language are:

(1) To reduce the direct cost of software development and maintenance
(2) To encourage the creation of more reliable systems
(3) To increase the mobility of programming staff
(4) To provide continuity of method and flexibility in choice of equipment by ensuring the portability of the application programs

Throughout the drafting of this standard the continued stability of RTL/2 has been a major aim. However, apart from changes to clarify or correct the original specification, minor alterations have been introduced which affect the scope of record component names, strings, LET definitions, and equality of label values. In addition the character set has been carefully designed to correct original features that, with hindsight, have proved to be inappropriate, yet without altering the semantics of valid programs.

It is virtually impossible to define precisely a high level language that enables programs to be executed with equal efficiency by all types of computer. Consequently, it has been necessary to leave certain areas of RTL/2 unspecified so that the characteristics of individual computers may be used to the best advantage. The most important areas are:

(a) Accuracy and range of real values
(b) Number of bits in an integer word
(c) Behaviour on arithmetic overflow

The second new standard in this issue is BS 5905 **Computer programming language CORAL 66,** (£14.50) which relates to a high-level language for applications on small to medium scale computers. It aims to achieve an overall economy, in terms of human effort, in the development of computer-based systems for real-time applications, and to protect user interests by encouraging the use of computers complying with this British Standard. Coral 66 provides a means of increasing the implementation efficiency of computer applications in environments where the input/output communication requirements may not have been standardized and are time-critical.

BS 5905 is based on the 'Official definition of CORAL 66' (published by HMSO in 1970), and aims to maintain the benefits of language stability by clarifying ambiguities in the original document while avoiding alterations or additions at the present time. The standard thus protects the large measure of standardization that already exists as a result of Ministry of Defence action in 1970 establishing CORAL 66 as part of official procurement policy, and which provides an authority that can be referred to in contractual arrangements.

INDUCTORS AND TRANSFORMERS

Advice on the preparation of technical documents for certain telecommunications components is provided in BS 5938 **Cores for inductors and transformers for telecommunications,** issued in the following two Parts:

Part 1 **Method of measurement** (£16.50) offers guidance on the drafting of those parts of specifications for magnetic cores that are concerned with methods of measuring the magnetic and electrical properties of cores made primarily of magnetic oxides or metallic powders and used in inductors and transformers for telecommunications and similar equipment. (This Part is technically identical with IEC 367–1 together with its supplementary documents 1A, 1B, 1C, 1D and 1E, issued by the International Electrotechnical Commission.)

Part 2 is entitled **Guide to the drafting of performance specifications** (£12.00) and is likewise concerned with inductor and transformer cores made of magnetic oxides or metallic powders and incorporated in telecommunications and similar equipment. Part 2 is identical with IEC 367–2 and its supplement 367–2A.

HI-FI PERFORMANCE

Minimum performance requirements for various types of domestic hi-fi audio apparatus are laid down in a new British Standard, BS 5942 **High fidelity audio equipment and systems: minimum performance requirements.** This is being published by BSI in several parts corresponding with international documents published by the International Electrotechnical Commission (IEC). The following three parts are now available:

Part 1 **General** (£3.60) gives the general conditions for measurement, requirements for interconnections, safety and published data. This part is identical with IEC 581–1.

Part 3 **Record playing equipment and cartridges** (£3.60) specifies the drive system in terms of speed deviation, wow and flutter, signal-to-rumble ratio and signal-to-hum ratio. This part also deals with the performance of the cartridge in terms of channel unbalance and separation, effective frequency range, tracking ability and angle, static vertical stylus force and stylus tip radius. Details of interconnections are provided as are the relevant methods of measurement. This part is identical with IEC 581–3.

Part 6 **Amplifiers** (£6.20) applies to linear and equalizing pre-amplifiers, power and integrated amplifiers, primarily intended for use with high quality reproducing aduio systems for home use. It deals with frequency response, harmonic distortion, output power, crosstalk, signal-to-noise ratio, thermal and electrical stability, short circuit protection, marking of controls and interconnections. This part is identical with IEC 581–6.

Other parts in preparation will cover tape recording and playback equipment, magnetic tape, radio. tuners, loudspeakers and headphones.

AUDIOMETERS

A new British Standard, BS 5966 **Audiometers** (£10.40) specifies requirements for audiometers designed primarily for use in determining hearing threshold levels by comparison with a chosen standard reference threshold level. Use of this standard should ensure that tests of human hearing, performed on the same ear but using different audiometers will give the same result under comparable conditions. Five categories of audiometers are classified in the standard according to their intended use. BS 5966 is identical with IEC Publication 645 and supersedes the existing BS 2980 which is now withdrawn.

INDUSTRIAL PROCESS MEASUREMENT AND CONTROL

Parts 1 and 2 of BS5967 **Operating conditions for industrial process measurement and control equipment** introduce a new multi-part series of standards providing a uniform listing and classification of various operating conditions. Intended to serve as a basis for the preparation of comprehensive specification by manufacturers and users of this equipment, the standard also aims to avoid problems that might result from failure to consider the effects of operating conditions on the performance of systems, or parts of systems.

Part 1 **Specification for temperature, humidity and barometric pressure** (£12.00) lists temperature gradient, humidity and barometric pressure conditions in specified locations, to which land-based and off-shore industrial-process measurement and control equipment may be exposed while being stored or transported, during periods when it is installed but inactive, or when actually operating. Part 1 is identical with IEC Publication 654–1.

Part 2 **Specification for power** (£9.00) gives the limiting values for power received by land-based and off-shore industrial-process measurement and control systems, or parts of systems. The influence quantities in this part are confined to those which may directly affect system performance, the effects on instrumentation being specifically excluded.

In both Parts, limit values are established for the operating conditions listed. Other operating conditions, including those for which characteristics are difficult to define and measure and for which adequate standards are not known to exist, will be covered in other Parts. A typical example is provided by corrosive atmospheres, which are difficult to classify due to the wide variety and concentration of corrosive substances and combinations that may be encountered.

No consideration is given in either Part to maintenance and repair conditions, or to operating conditions directly related to fire and explosion hazards or to nuclear radiation.

LABORATORY RESISTORS

BS 6011 **Laboratory a.c. resistors** (£7.60) specifies requirements for resistors of this type used for frequencies up to 100 kHz. A new standard, it deals with classification, general requirements, permissible variations of resistance with frequency, electrical and mechanical requirements, markings and symbols. It is identical with IEC Publication 477–2.

PACKAGING OF ELECTRONIC COMPONENTS

The first Part of a new British Standard of particular interest to the electronic equipment manufacturing industry has been issued as BS 6062 **Packaging of electronic components for automatic handling** Part 1 **Specification for tape packaging of components with axial leads on continuous tapes.** (£5.00)

Part 1 specifies the correct spacing, lead length, orientation and alignment necessary for operations involving automatic component insertion machines. It has been prepared in order to standardize the methods, general dimensions and preferred tolerances for the taping of such components as resistors, capacitors, diodes etc. installed in telecommunication equipment and other electronic devices employing similar assembly techniques.

The new standard is identical with IEC Publication 286–1, published by the International Electrotechnical Commission (IEC). Further Parts are to be published in due course and will cover multiple and other terminations and leadless components.

RADIO EQUIPMENT FOR SATELLITE EARTH STATIONS

The British Standards Institution has published the first two parts of BS 6080 **Measurement for radio equipment used in satellite earth stations**, which aims to standardize methods of measuring characteristics specified in the customer/manufacturer contracts and facilitate comparison of results by different observers on different equipment.

The first of these documents is Part 1 **General** (£9.00) which defines conditions for 'type' and 'acceptance' tests and gives details of selected methods of making measurements for assessing the essential properties of satellite earth stations and associated equipment. Part 1 is identical with IEC 510.

The second new standard is Part 2.1 **Measurements for sub-systems** (£16.50), which covers general requirements for methods of measuring the electrical characteristics of sub-systems used in earth station equipment for communication through orbiting satellites. A sub-system is a combination of circuits or devices performing a given function such as modulation, frequency-conversion, amplification etc. Methods of measuring the electrical characteristics of antennas are also dealt with in this Part. Part 2.1 is identical with IEC 510–2–1.

ELECTRICAL AND ELECTRONICS SYMBOLS

A new training package devised by the British Standards Institution will be of considerable benefit to teachers and students of electrical and electronics engineering and associated disciplines. Entitled PD 7303 **Electrical and electronic graphical symbols for schools and colleges** (Price £6.00 plus VAT. There is no discount for BSI subscribing members). It aims to reduce the expense and inconvenience that could otherwise be involved in consulting the many different sections that make up the major reference source on the subject (ie BS 3939 **Graphical symbols for electric power, telecommunications and electronics diagrams**).

PD 7303 comprises an A4-size booklet which contains 270 of the more commonly used symbols, their official descriptions (and popular alternatives in some cases), explanatory notes, and an A1-size wall chart displaying a selection of 150 symbols with suitably condensed descriptions. All the symbols are identified by numbers and indexed so that the user may refer to the appropriate section of BS 3939.

The publication is designed to allow the pages to be readily used as master copies for duplicated hand-outs of symbols, either in the form of groups or as individual sections (eg semi-conductor devices and logic elements). BSI has granted permission for material to be duplicated for use by any single class within the one educational establishment, and this facility is allowed for in the cost of the package.

TEXT COMMUNICATION REPORT

Prompt action is vital to remedy the present lack of standardization in an important branch of UK industry. This warning is given in a special report entitled **The need for and role of, standards in text communication,** (£2.50) which contains the first recommendations of a BSI working party, set up in 1979 by committee DPS/OMS/2 Text and facsimile transmission, to examine the implications of the rapidly-expanding use of word-processing equipment, particularly for communication purposes.

DPS/OMS/2 is responsible for integrating the activities of both office equipment and data processing committees, and the report stresses that communication of text in a form that will meet future user requirements will involve the interworking of many different (often proprietary) types of equipment. Efficient working will require a much stricter adherence to national standards, preferably based on international standards, than is found in current computer-linked equipment. Yet, at present, much of the documentation considered necessary for the communication of text does not exist nor, in some areas, is any work being done.

CAPABILITY APPROVAL

A new information document about capability approval under the BS 9000 System covering electronic components of assessed quality is now available. The pamphlet contains a concise explanation of the need for capability approval and the benefits it offers to users of custom-made printed circuits, integral circuits, transformers and inductors. It is hoped that the document will meet the growing need for information about this increasingly important aspect of the BS 9000 System.

Copies of this pamphlet are available from the Certification and Assessment Department, BSI, Maylands Avenue, Hemel Hempstead, Herts, HP2 4SQ (Tel: 0442 3111 ext 303). Free of charge for small quantities, £15 per hundred for bulk supplies.

NEW BSI INFORMATION TOOL

An easy-to-use thesaurus is essential to the smooth running of any modern technical information system. **The BSI ROOT Thesaurus** (£155; updating to March 1982, £50) was originally designed to assist BSI and its Technical Help to Exporters Department to establish a 'databank' of British and other major Standards and regulations from all over the world. All the main branches of engineering and science are covered, with particular emphasis on measurement, testing and instruments, environmental and safety engineering. Supporting fields such as social sciences and administration are included.

ROOT was developed in parallel with an English/French thesaurus for the ISONET database exchange system used by the International Organization for Standardization (ISO). Translations into additional languages (already in preparation within ISONET) may later be incorporated for multilingual applications.

ARE YOU PLUGGING IN THE RIGHT MARKET?

Despite moves toward an internationally harmonized system of plug and socket outlets, the types of plugs used throughout the world remain as varied as ever. For this reason the Technical Help to Exporters service of the British Standards Institution publishes an international survey of electrical plugs, and the third edition of this useful document is now available.

Entitled **Electrical plugs—an international survey** (£25.00), it identifies the various types of plugs in general use overseas for single phase domestic equipment. The previous edition has been completely revised and expanded to give up-to-date information on no less than 148 countries. The latest developments on an internationally harmonized system are also detailed in the new publication. The survey enables manufacturers and exporters of electrical equipment to identify quickly the types of plugs in common use throughout the world.

Technical Help to Exporters, Maylands Avenue, Hemel Hempstead, Herts HP2 4SQ. Telephone: 0442 3111.

The Tenth Clerk Maxwell Memorial Lecture

The Complete Physicist: James Clerk Maxwell, 1831–1879

Professor R. V. Jones, C.B., C.B.E., F.R.S.*

Given at the University of Leeds on 7th July 1981

Both in pure science and in its applications, the twentieth century has seen tremendous advances; and in our pride in the achievements of our own generation, and in our enthusiasm for what may be done by the generation to come, it is fitting to remember that these achievements have only been possible because, as Isaac Newton so humbly said of himself, we have stood on the shoulders of giants. I am therefore much honoured by your invitation to recall the memory of one of the greatest of all these giants, James Clerk Maxwell, to mark the one-hundred-and-fiftieth anniversary of his birth. His many contributions ranged over, and indeed widened, the entire range of physics of his day; and it was his deep insight and imagination that led to the foundation of radio science and engineering which form the special interest of your Institution.

Early Life

As for his family name, this was originally Clerk; his father, John, added Maxwell on inheriting the estate of a Miss Maxwell near Castle Douglas in Kirkcudbrightshire: and it is significant that despite his complete lack of pomposity James generally used both surnames in his signature. His mother, who before her marriage had been Frances Cay, came from Northumberland. He was born at the family house at 14 India Street, Edinburgh, on 13th June 1831, and spent a happy childhood on the Kirkcudbrightshire estate of Glenlair until the untimely death of his mother in 1839. His father, a lawyer, was faced with the problem of educating and looking after him, and so brought in a tutor; the latter turned out to be a bully, and James made at least one attempt to escape from him by sailing out in a washtub on the duck pond. James's biographers[1] are convinced that the bullying tutor was the cause of his later hesitation of manner and obliquity of reply. His father realized that things were going wrong, and therefore decided in 1841 to send him to Edinburgh Academy, at which he arrived in serviceable but unconventional clothing designed by his too-rational father. He returned from his first day at school to his new home (with his aunt Mrs Wedderburn at 31 Heriot Row) with his tunic in rags, his skirt lost, and his frilled collar rumpled and torn. 'Daftie' was the nickname the other boys gave him and it stuck to him throughout his schooldays.

But before he left the Academy to go to Edinburgh University, James had read a paper,[2] 'On the Description of Oval Curves and those having a Plurality of Foci', to the Royal Society of Edinburgh. He was fourteen years old at the time, and the paper described an original and simple method of constructing Cartesian Ovals—curves which had been specified by Descartes, and which are of some interest in the design of optical instruments, since lenses with such surfaces could be made free of spherical aberration. This youthful paper gave an earnest of his great powers of geometrical thinking, and of the contributions that he was to make to the theory of optical instruments.

Undergraduate Years

When he went to Edinburgh University in 1847, he studied Mathematics, Natural Philosophy, Chemistry and Mental Philosophy. In these courses he came principally under the influence of two men, J. D. Forbes in Natural Philosophy, and William Hamilton in Mental Philosophy. The latter was one of the principal figures in the school of Scottish commonsense philosophy, which had originated with the 18th century Philosophical Society of Aberdeen whose most influential member was Thomas Reid, Professor of Philosophy at King's College. The Aberdeen philosophers, in their attempts to defend moderate Presbyterianism against the Rationalist challenge of David Hume, had encountered the difficulty that many of us currently face in battling with the prospect of anarchy based on specious rationalism: the latter can lead to conclusions which in Reid's[3] words contradict 'certain principles which the constitution of our nature leads us to believe, and which we are under a necessity to take for granted in the common concerns of life'. These are moral principles which can be no more proved than the axioms of Euclid and yet which are correspondingly essential as the foundations for a framework of morality.

This seems to be a remote base from which to trace the rise of pictorial thinking in British Physics, of which Maxwell was to be the outstanding exponent, but it is what Dr Richard Olson set out to do in his book 'Scottish Philosophy and British Physics 1750–1880' (Princeton University Press, 1975). Perhaps the most

* *Department of Natural Philosophy, Aberdeen University.*

0033–7722/81/110527+13 $1.50/0

Clerk Maxwell as a young man (from a portrait at Trinity College, Cambridge, by permission of the Master and Fellows)

suggestive argument is that with its reliance on commonsense plus a few moral axioms, Scottish philosophy had greater faith in the pure 'geometry of visibles', proving results by Euclidian arguments, than in analysis, where the algebra of infinite series was known to be dangerous, and where some of the entities such as the square root of -1 had no commonsense physical interpretation. Certainly, there was strong interaction in the Scottish universities between the Moral and the Natural philosophers, and undergraduates were exposed to both influences in the traditional Scottish degree.

Be all that as it may, Maxwell developed deep interests both in Natural Philosophy and in Mental Philosophy, at the same time viewing his professors both with the affection and with the humour that were his lifelong characteristics. Writing[4] to his friend and later biographer Lewis Campbell in 1848, he described one of Forbes's demonstrations: 'On Saturday the natural philosophers ran up Arthur's Seat with the barometer. The Professor set it up at the top and let us pant at it till it ran down with drops. He had not set it straight and made the hill grow 50 feet; but we got it down again.'

In 1850 it was decided that Maxwell should go to Cambridge, after some delay through indecision arising from his father's original wish for him to become an advocate like himself. James, however, decided that it was, as he said, 'another kind of laws' that he was called upon to study and, with his father's approval, he went to Peterhouse in 1850. The next year he transferred to Trinity because he would stand a better chance of a Fellowship; but he remained under the guidance of Hopkins of Peterhouse, who had also been the tutor of Kelvin and Stokes. Hopkins said of Maxwell that it seemed impossible for him to think incorrectly on physical subjects, but that in his analysis he seemed far more deficient.

His contemporaries noticed that whenever the subject admitted he had recourse to diagrams, even though his fellow students might more easily solve them by analysis. Once he surprised both them and the lecturer, when the latter had filled the blackboard three times over in solving a problem in analytical geometry, by pointing out that with a figure and a few lines of purely geometrical reasoning the answer would come out readily. He was subsequently to write many geometrical papers including, for example, a delightful study[5] in topology 'On Hills and Dales' (1870), which contained such results as 'number of summits = number of passes plus 1', and the whimsical use of 'immit' as the opposite of 'summit'.

Post-Graduate Research and Writings

Dr Robert Walker, a pupil of Maxwell at Aberdeen, wrote of his lectures: 'He was always most lucid when he fell on geometrical methods'. This was typical of his attitude to physical problems, which led him more to proceed from one distinct idea to another than to trust to symbols and equations, at least when exploring new territory. This approach was exemplified by a paper[6] which he wrote in 1855, shortly after graduation, on Faraday's Lines of Force:

> 'The first process therefore in the effectual study of the science, must be one of simplification and reduction of the results of previous investigation to a form in which the mind can grasp them. The results of this simplification may take the form of a purely mathematical formula or of a physical hypothesis. In the first case we entirely lose sight of the phenomena to be explained; and though we may trace out the consequences of given laws, we can never obtain more extended views of the connexions of the subject. If, on the other hand, we adopt a physical hypothesis, we see the phenomena only through a medium, and are liable to that blindness to facts and rashness in assumption which a partial explanation encourages. We must therefore discover some method of investigation which allows the mind at every step to lay hold of a clear physical conception, without being committed to any theory founded on the physical science from which that conception is borrowed, so that it is neither drawn aside from the subject in pursuit of analytical subtleties, nor carried beyond the truth by a favourite hypothesis.'

Although, therefore, valuing the physical approach based on analogical reasoning, Maxwell was alive to the danger of carrying analogy too far:

> 'In order to obtain physical ideas without adopting a physical theory we must make ourselves familiar with the existence of physical analogies. By a physical analogy I mean that partial similarity between the laws of one science and those of another which makes each of them illustrate the other. Thus all the mathematical sciences are founded on relations between physical laws and laws of numbers, so that the aim of exact science is to reduce the problems of nature to the determination of quantities by operations with numbers. Passing from the most universal of all analogies to

a very partial one, we find the same resemblance in mathematical form between two different phenomena giving rise to a physical theory of light.

'The changes of direction which light undergoes in passing from one medium to another, are identical with the deviations of the path of a particle in moving through a narrow space in which intense forces act. This analogy, which extends only to the direction, and not to the velocity of motion, was long believed to be the true explanation of the refraction of light; and we still find it useful in the solution of certain problems, in which we employ it without danger, as an artificial method. The other analogy, between light and the vibrations of an elastic medium, extends much farther, but, though its importance and fruitfulness cannot be over-estimated, we must recollect that it is founded only on a resemblance *in form* between the laws of light and those of vibrations. By stripping it of its physical dress and reducing it to a theory of "transverse alternations", we might obtain a system of truth strictly founded on observation, but probably deficient both in the vividness of its conceptions and the fertility of its method.'

The foregoing passage shows that from the beginning of his research career Maxwell had a profound understanding of the function of analogy in science. He could write humorously as well as seriously on this subject, as he showed in an essay[7] of February 1856 (which can be found in his biography), in which he asked the question 'Are there Real Analogies in Nature?' It starts with a characteristic touch:

'Now, as in a pun two truths lie hid under one expression, so in any analogy one truth is discovered under two expressions. Every question concerning analogies is therefore the reciprocal of a question concerning puns, and the solutions can be transposed by reciprocation. But since we are still in doubt as to the legitimacy of reasoning by analogy, and as reasoning even by paradox has been pronounced less heinous than reasoning by puns, we must adopt the direct method with respect to analogy and then, if necessary, deduce by reciprocation the theory of puns'.

One point which Maxwell made regarding the analogies associated with light deserves special notice. The argument whether light was to be thought of as the motion of waves or of particles had been going on for nearly two centuries, and as recently as 1850 the question had appeared to be settled in a complete triumph for the wave theory by the experiments of Foucault and Fizeau,[8] who had shown that light travelled more slowly in water than in air. But as is evident from what Maxwell said, he was not stampeded into confounding the wave analogy with total reality, but instead warned his readers that the vibration of an elastic medium was itself only an analogy as far as light was concerned; and we can imagine that he would therefore not have been so surprised as would have been many of his contemporaries when it was much later discovered that light had the particulate properties described as photons, in addition to the wave properties that in 1850 appeared exclusively paramount.

What Maxwell set out to do in his 1855 paper was to take Faraday's pictorial ideas of Lines of Force, and to give them analytical expression. Faraday had been diffident about his ideas, and especially about his conjecture that light itself might be some kind of wave propagated along his Lines of Force. He had been led to propound them in an impromptu discourse that he had given at the Royal Institution in 1846 when the official lecturer for the evening, Charles Wheatstone, had taken fright and absconded just before the discourse was to have started. Apologizing for their publication, Faraday[9] wrote 'I do not think that I should have allowed these notions to escape from me, had I not been led unawares, and without previous consideration, by the circumstances of the evening at which I had to appear suddenly and occupy the place of another'. But the train of thought thus started by Faraday was picked up by Maxwell, in whose mind it must have simmered through the changes that were shortly to occur in his life. Maxwell certainly believed that his mind operated in this way for, as he wrote[10] to a friend, 'I believe there is a department of mind conducted independent of consciousness where things are fermented and decocted, so that when they are run off they come clear'.

Another of his interests in his postgraduate days at Cambridge was colour vision, and in a paper[11] later published in 1857 he proposed what is now known as the additive system of colour photography, such as is used in colour television.

'Let a plate of red glass be placed before the camera, and an impression taken. The positive of this will be transparent wherever the red light has been abundant in the landscape, and opaque where it has been wanting. Let it now be put in a magic lantern, along with the red glass, and a red picture will be thrown on the screen.

'Let this operation be repeated with a green and a violet glass, and, by means of three magic lanterns, let the three images be superimposed on the screen. The colour of any point of the screen will then depend on that of the corresponding point of the landscape; and, by properly adjusting the intensities of the lights etc., a complete copy of the landscape, as far as visible colour is concerned, will be thrown on the screen.'

At the same time he was writing more general essays, spurred by his deep interest in scientific method, and there were several other delightful examples from Maxwell's pen at this period. A month after the essay on puns he was asking whether autobiography was possible, remarking that 'When a man once begins to make a theory of himself, he generally succeeds in making himself into a theory'.

A First University Chair at Aberdeen

By 1856 Maxwell had already been elected a Fellow of his College in Cambridge; but just as the papers on Faraday's Lines of Force and colour vision were appearing, he decided to apply for the Chair of Natural Philosophy at Marischal College, Aberdeen. One of the attractions of this new position was that he would be nearer to his father, with whom he had much in common and to whom he wrote:

'As far as writing Testimonials goes, there is a good deal, and if you believe the Testimonials you would think the Government had in their hands the triumph or downfall of educaton generally, according as they elected me or not.'

'I had a letter from Dr Swan of Edinburgh, who is a candidate, asking me for my opinion, which I gave him, so far as I had one'.

Unhappily, John Clerk Maxwell died shortly before the Aberdeen election took place; but James let his application stand because it had his father's warm approval. He was appointed at the end of April 1856 and took up duty in the following November.

He prepared both to take his undergraduate classes and to introduce himself to his colleagues through his Inaugural Lecture. Of this he wrote[12] (14th October 1856): 'Now I am writing a solemn address or manifesto to the Natural Philosophers of the North, which I am afraid I must reinforce with coffee and anchovies, and roaring hot fire and spread coat-tails to make it all natural.' The audience included most of the Senatus Academicus, 'a very large number of students' and Piazzi Smith, the Astronomer Royal for Scotland (one of the very few Fellows, incidentally, ever to resign from the Royal Society of London, after which he styled himself 'Charles Piazzi Smith, Late F.R.S.'). Maxwell's manuscript of his lecture still exists in the University Library at Cambridge. It has been reproduced in a paper[13] that I wrote for *Notes and Records of the Royal Society* in 1973.

In it Maxwell told his students:

> 'My duty is to give you the requisite foundation and to allow your thoughts to arrange themselves freely. It is best that every man should be settled in his own mind, and not be led into other men's ways of thinking under the pretence of studying science. By a careful and diligent study of natural laws I trust that we shall at least escape the dangers of vague and desultory modes of thought and acquire a habit of healthy and vigorous thinking which will enable us to recognize error in all the popular forms in which it appears and to seize and hold fast whether it be old or new.'

And he emphasized the need for experiment:

> 'I have no reason to believe that the human intellect is able to weave a system of physics out of its own resources without experimental labour. Whenever the attempt has been made it has resulted in an unnatural and self-contradictory mass of rubbish!'

Maxwell devoted much of his effort to teaching his classes, which were in the main composed of elementary students, taking care to illustrate his lectures with as many practical demonstrations as possible. Further, his enthusiasm for education led him to continue at Aberdeen the practice that he had started at Cambridge of lecturing to working men in the evenings, and it is hard to believe that anyone with such evident enthusiasm could be anything but a good teacher. Yet Maxwell's ability as a teacher has often been questioned. It is true that his own ability and that of his students were very different, and the richness of his imagination and the swiftness of his wit may have been hard for an elementary class to follow. It is true that, as his Cambridge pupil Sir Horace Lamb said:[14]

> 'He had his full share of misfortunes with the blackboard, and one gathered the impression, which is confirmed I think by his writings, that though he had a firm grasp of essentials, and could formulate great mathematical conceptions, he was not very expert in the details of minute calculations. His physical instinct saved him from really vital errors.'

But Lamb also said of Maxwell's lectures:

> 'They had a great interest and charm for some of us, not so much for the sake of the subject matter, which was elementary, as in the illuminating glimpses we got of the lecturer's own way of looking at things, his constant recourse to fundamentals, and even his expedients when in a difficulty, the humorous and unpremeditated digressions, the occasional satirical remarks, and often a literary or even poetical allusion.'

Maxwell found, though, that his humorous digressions were not always appreciated at Aberdeen, where shortly after the beginning of his second session he wrote 'No jokes of any kind are understood here. I have not made one for two months, and if I feel one coming I shall bite my tongue'.[15]

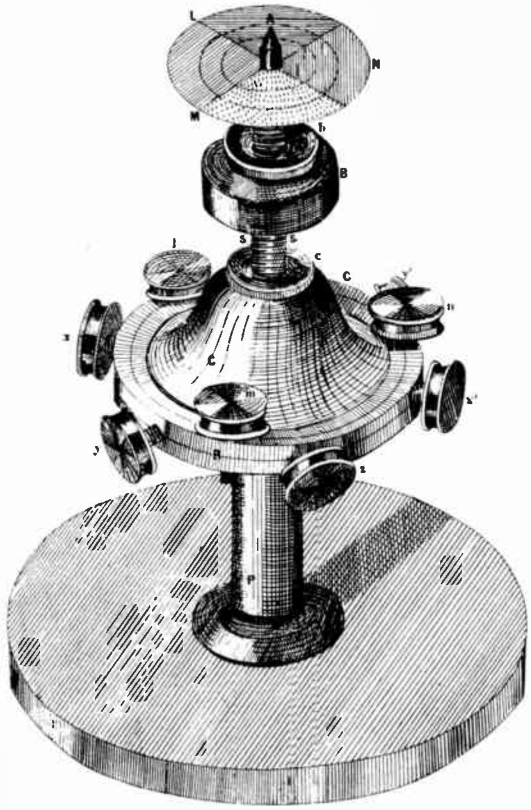
Work on Dynamics and Optics

His first serious research at Aberdeen was a detailed theoretical investigation of the structure of Saturn's Rings, which had been set as the subject for the Adams Prize of 1857. He came to the conclusion that the Rings cannot be solid and uniform, since they would wobble and crash into the planet. Nor could they be a continuous fluid, and the only possible structure was a swarm of small particles, which he showed could be stable, with the prediction that the particles on the inside of a Ring must orbit more rapidly than those on the outside. Not only was this subsequently confirmed by Doppler Effect observations, but recently the spectacular mission of *Voyager I* has produced a wealth of evidence by close inspection.

His theory of Saturn's Rings was illustrated by a mechanical model, constructed to his design by the Aberdeen firm of Smith and Ramage. Maxwell's outstanding work as a theoretical physicist has often caused his experimental ability to be overlooked. His letters[16] from Aberdeen about the Ring model show how close he was to the practical side: 'I am happy in the knowledge of a good instrument maker, in addition to a smith, an optician and a carpenter' (22nd October 1857).

Another piece of Aberdeen craftsmanship was Maxwell's Dynamical Top, which he demonstrated to the Royal Society of Edinburgh on 20th April 1857, introducing it as 'The top which I have the honour to spin before the Society.....' He took the top with him on a visit to Cambridge in 1857, and demonstrated it one evening to his friends. It was spinning when they left. Next morning Maxwell saw some of them returning to his rooms, jumped out of bed and set the top spinning, and pretended still to be sleeping when they came in. The phenomenal properties of Maxwell's top became a Cambridge legend.

Maxwell not only appreciated craftsmanship but himself delighted in using his hands, starting with his childhood ability in sketching. At one stage he became interested in the design of the zoetrope, a forerunner of the cinematograph in which a quick succession of silhouetted figures, each in a slightly different attitude from its predecessor, could be projected into a field of view and thus give the impression that the figure was



Maxwell's dynamical top. The top was used to spin the sectored disk for Maxwell's experiments on colour mixing.

running or dancing. His sense of mischief led him to combine this device with a demonstration of a serious optical phenomenon that he was about to demonstrate to Sir William Thomson, later Lord Kelvin. Maxwell invited him to look through the eyepiece to observe the phenomenon, where to Kelvin's surprise he saw the phenomenon that Maxwell had indeed described, but with a little man dancing around at the same time. Puzzled, Kelvin asked Maxwell 'What is the little man there for?' 'Have another look', Maxwell replied, 'you ought to be able to see what he is there for!' No wiser after a second look, Kelvin impatiently said 'Well, what is he there for?' 'Just for fun, Thomson!' said Maxwell.

Maxwell's skill with his hands also extended to optical polishing, and years later—when he was at Cambridge—he succeeded in cutting a piece of crystal to show the phenomenon known as conical refraction. Delighted with his success, Maxwell happened to meet Todhunter, the famous mathematical tutor, quite possibly during the afternoon perambulation on King's Parade immortalized by F. M. Cornford in 'Microcosmographia Academica.' Maxwell invited Todhunter to come back and witness conical refraction for the first time, to meet with the surprising reply 'No, I have been teaching it all my life, and I don't want all my ideas upset by seeing it now!'[17]

During Maxwell's first year in Aberdeen, the idea was mooted of inviting the British Association to meet there in 1859. On 28th August 1857 there was a public meeting of interested citizens, who decided that they had no suitable hall for the main assembly. They therefore raised the money themselves by personal subscription, and the building, now the Music Hall, was built and ready within two years. Maxwell was one of those who subscribed, with the result that for many years after his death there were from time to time small dividends to be paid by a firm of advocates. It was this that resulted in an astonishing advertisement in one of the Aberdeen newspapers of the 1920s asking anyone who knew of the present whereabouts of Mr James Clerk Maxwell to get in touch with the firm.

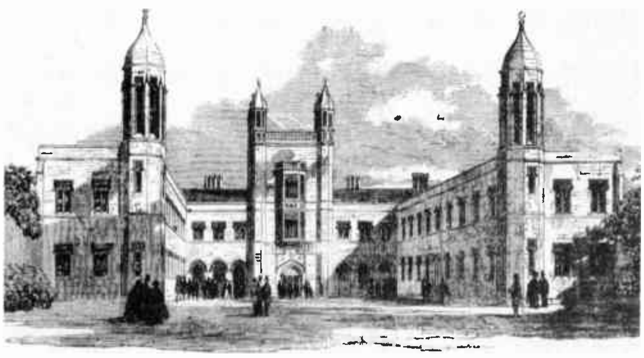
I heard the story from an old school inspector about thirty years ago, who had throughout his life collected stories of Maxwell by talking to men who had been among his students. (One old farmer, for example, could just remember Maxwell as the professor who had stood one of the students on an insulating stand and had then 'pumpit him fu' of electricity' so that his hair stood on end). The inspector, seeing the advertisement, went along to the advocates and was greeted with some interest. Before he answered their question he asked the head of the firm 'Do you really mean to tell me that you have never heard of James Clerk Maxwell, the most famous man who ever walked the streets of Aberdeen?' 'No', replied the advocate 'who was he?' The advocate listened with interest as the inspector outlined some of Maxwell's achievements and then said 'This is all very interesting. Now I will tell you why we put the advertisement in the paper. For years we have been sending dividends arising from the Music Hall addressed to "Mr James Clerk Maxwell, Marischal College", but we have always had them returned "Not known".'

In the list of participants in the British Association meeting, Maxwell's address is given as 13 Victoria Street, which happened also to be the address of Principal Dewar of Marischal College. In the meantime Maxwell had met Principal Dewar's daughter Katherine, and they were married in June 1858. Her first record[18] in Maxwell's work occurs in one of the papers that he was writing on colour vision, where he measures the colour sensitivity of the eyes of different observers, 'J' obviously being himself and 'K' his future wife (whom he later described[19] as 'a very accurate observer of the normal type'.) He broke the news of his impending marriage to his aunt, Miss Cay:

> Dear Aunt,
> This comes to tell you that I am going to have a wife. Don't be afraid; she is not mathematical, but there are other things besides that, and she certainly won't stop the mathematics. The only one that can speak as an eye-witness is Johnnie, and he only saw her when we were trying to act the indifferent. We have been trying it since, but it would not do, and it was not good for either. So now you know who it is, even Katherine Mary Dewar (hitherto).

The most important paper that Maxwell wrote during his time at Aberdeen was one that he read at the British Association meeting under the title 'Illustrations of the Dynamical Theory of Gases.'[21] It is one of the historic papers of physics, for in an elegantly simple—if non-rigorous way he solved the problem of how the velocities of the colliding molecules in a volume of gas would distribute themselves, the solution ever since being known as The Maxwell Distribution Law.

Shortly after the paper was published in 1860, unhappily, Aberdeen dispensed with Maxwell's services. The basic reason was that the two separate degree-giving colleges of King's and Marischal were fused into the present University, and so there were now twice as many professors as were needed. The general rule was that the junior professor was to be retained, and the senior retired. The reasoning behind this rather curious rule may have been that since the retired professor was to be paid a pension, on the average the senior man would have the shorter time to live and therefore be the less expensive. In one case alone, however, was the rule reversed: this was Natural Philosophy, where the senior professor, David Thomson, of King's was retained and the junior, Maxwell, retired. Thomson had done much to promote the fusion of the colleges and had a good reputation as a teacher, although his research record was negligible. One assessment of the manoeuvres behind



Marischal College, Aberdeen, drawn for *The Illustrated London News*, 1859. Maxwell's lecture room is said to have been on the first floor behind the tower to the left of the drawing.

Maxwell's retirement was given by Dr Robert Walker in the *Aberdeen University Review* for June 1916, suggesting Thomson's nickname of 'Crafty' was by no means undeserved.[22] Another version was given me by the late Sir Edmund Whittaker who was told by Mrs P. G. Tait, the wife of Maxwell's great friend, that Mrs Maxwell herself was partly responsible in that she much favoured the prospect of being the wife of a country gentleman, as she could be if Maxwell were to retire to Glenlair.

The Move to King's College, London

Whatever the explanation behind the move, the Maxwells left Aberdeen in 1860, but not to spend their entire time at Glenlair, for he first applied for the vacant Chair of Natural Philosophy at Edinburgh, where P. G. Tait was preferred, and then for the Chair at King's College London, which he took up towards the end of the year. His Inaugural Lecture at King's College, like that at Aberdeen, still exists in manuscript and is now in print[23] It is indeed very similar to the Aberdeen lecture but it includes a passage that reflects his growing interest in the History of Science which should give would-be historians of science something to think about:

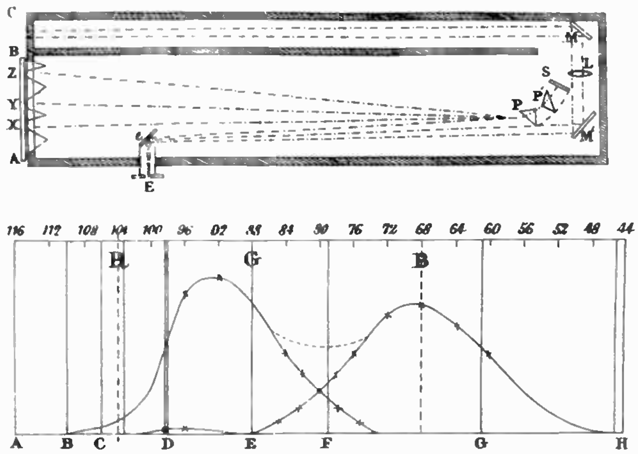
> 'We cannot understand the steps by which the human mind has advanced to its present state of knowledge, till we ourselves have some experience of what that state of knowledge is. When we have encountered and overcome the resistance of our minds to the acquisition of new ideas, we shall be the better able to appreciate the labours of those who, for the first time thought out those ideas, and transmitted them to us for a perpetual possession'.

This raises the question of whether one can genuinely appreciate the history of science without oneself having struggled to advance it.

In 1861 Maxwell was able to show a practical result for the ideas on colour vision which he had explored in Aberdeen and which he had anticipated at Cambridge, for he was able to demonstrate in a Royal Institution discourse the method of colour photography that he had proposed by the superposition of images obtained in red, green, and blue lights. The photograph was of a tartan bow, and the negatives still exist in the Cavendish Laboratory at Cambridge, along with several of his instruments, most of them made in Aberdeen. No doubt Maxwell had paid for them out of his own pocket and so was not inclined to leave them behind when he departed. The Research Department of Kodak has made prints from the negatives, and so we can still see what he showed his audience at the Royal Institution. To some extent the result is rather misleading in that Maxwell's photographic plate could hardly have had enough sensitivity in the red to produce an appreciable image, but it happened that his red filter also transmitted a little at the other end of the spectrum, and so an image appeared on the red plate although it was due to light in the violet or ultraviolet. The result may truly be claimed as the first colour photograph, even though the rendering is partly false, and it pointed the way for all subsequent additive methods of colour photography, and is basically the method used in colour television. For his work on colour vision the Royal Society of London awarded Maxwell its Rumford Medal—its premier distinction for work in physics—in 1860.

The Electromagnetic Theory of Light

An even more important development which had been simmering in Maxwell's mind now began to come clear. Developing his ideas regarding Faraday's Lines of Force and Ray Vibrations, he tried to make a mental model of the processes that might be going on in the phenomenon of electromagnetic induction. He found it necessary to suppose that each cell of the medium which transmitted magnetic action 'possesses elasticity of figure, similar in kind, though different in degree, to that observed in solid bodies'. This model, along with the concept of a medium in a state of electric tension, led him in 1862 to conclude that even an insulating dielectric would experience the phenomena associated with an electric current while the



Maxwell's 'colour box', designed to mix three spectral colours from slits at X, Y and Z by dispersion through the prisms P and P. The light is reflected back through the prisms by the mirror S to the eyepiece E.

tension was being built up. Since this was probably the first mention of his famous displacement current, it is worth quoting what he said:[24]

> 'In a dielectric under induction, we may conceive that the electricity in each molecule is so displaced that one side is rendered positively, and the other negatively electrical, but that the electricity remains entirely connected with the molecule, and does not pass from one molecule to another.
>
> 'The effect of this action on the whole dielectric mass is to produce a general displacement of the electricity in a certain direction. This displacement does not amount to a current, because when it has attained a certain value it remains constant, but it is the commencement of a current, and its variations constitute currents in the positive or negative direction, according as the displacement is increasing or diminishing.'

Maxwell's conception of the displacement current therefore came from thinking along purely physical lines and not, as is sometimes related in text books, by consideration of a missing term in Ampère's circuital relation.[25]

His next step was to perceive an analogy between electrical displacement in a dielectric and elastic displacement in a solid medium. In other words, there was a kind of elasticity, or restoring force, associated with displacement in a dielectric, and it was well known that a medium that had both elasticity and inertia could transmit waves. Moreover, Maxwell believed that even empty space was filled with the mysterious medium known as the aether, and he accepted from Kelvin the idea that this mysterious medium differed only from tangible media in degree rather than kind, and that it was of 'small but real density'. He proceeded to work out the speed of waves that the aether should transmit, and found that they came out to be very near the velocity of light. He began by saying:[26]

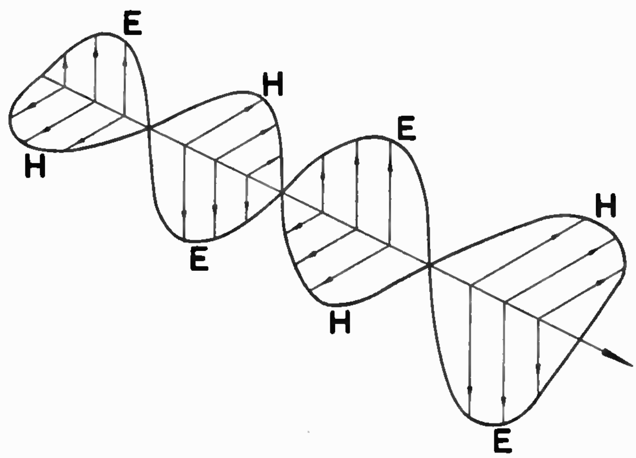
> 'In the following investigation I have considered the relation between the displacement and the force producing it........I have deduced from this result the relation between the statical and dynamical measures of electricity, and have shewn, by a comparison of the electro-magnetic experiments of MM. Kohlrausch and Weber with the velocity of light as found by M. Fizeau, that the elasticity of the magnetic medium in air is the same as that of the luminiferous medium, if these two coexistent, coextensive, and equally elastic media are not rather one medium......'

And he strengthened his statement eight pages later:

> 'The velocity of transverse undulations in our hypothetical medium, calculated from the electro-magnetic experiments of MM. Kohlrausch and Weber, agrees so exactly with the velocity of light calculated from the optical experiments of M. Fizeau, that we can scarcely avoid the inference that *light consists in the transverse undulations of the same medium which is the cause of electric and magnetic phenomena.*'

So here, in these papers of 1861 and 1862, we have the first conception of the displacement current, and the first clear statement that light is an electromagnetic phenomenon. Maxwell's use of italics may indicate his delight.

In this work, Maxwell had used a mechanical model to simulate the properties of the aether, and he followed it by attempting to construct a theory that would be more general than the model which had led to such a fascinating conclusion. Papers published by him in 1864 took the subject substantially further, and they made clear that he considered that 'so-called vacua' are filled with the mysterious medium known as the aether, and that even in the emptiest space 'there is always enough of matter left to transmit the undulations of light and heat' and he went on to quote Kelvin's work on the density of the aether: 'We may therefore receive, as a datum derived from a branch of science independent of that with which we have to deal, the existence of a pervading medium, of small but real density.' Since the medium had both mass and elasticity, it could store both kinetic and potential energy, and could also be the seat of the displacement current, just as could a tangible dielectric. He proceeded to conclude that the medium 'having such a constitution may be capable of other kinds of motion and displacement than those which produce the



Representation of Maxwell's mental picture of a electromagnetic wave with transverse electric and magnetic vibrations at rightangles to the direction of propagation.

phenomena of light and heat, and some of these may be of such a kind that they may be evidenced to our senses by the phenomena they produce'.[27] This is effectively the point at which Maxwell first speculates about electromagnetic phenomena with characteristics other than those of light and heat, and which may therefore be said to be the first pointer to the possible existence of radio waves.

In parallel with this work, Maxwell was also pursuing the dynamical theory of gases, based on the concept that gases consist of myriads of molecules flying in all directions and colliding with one another and the walls of their container. His detailed analysis suggested the astonishing conclusion that the viscosity of a gas should, over a very wide range of pressures, be independent of the pressure, when commonsense would have suggested that when a gas is compressed and therefore becomes more dense, it would offer more viscous resistance to a body moving through it. Maxwell made experiments to check the theory, and found that his surprising predicton was in fact fulfilled.[28] It was a triumph which resulted in the widespread acceptance of the kinetic theory and the existence of molecules and it so delighted Lord Rayleigh that he wrote that physics contained 'no more beautiful or telling discovery than that gaseous viscosity is the same at all densities'.[29]

Yet another fundamental contribution by Maxwell came during this same period, when he became concerned with establishing the precise values of the basic units of electricity, since he had shown that they were related to the velocity of light. In one of these experiments a coil had to be rotated at a constant speed, and so Maxwell used an adaptation of the governor that James Watt had developed for his steam engine. Maxwell thus became interested in the theory of governors, and in 1868 wrote a paper 'On Governors,'[30] of which relatively little notice was taken until the end of World War II, when automatic control of aircraft and industrial mechanisms became important, and it was realized that the basic theory was already available in Maxwell's paper.

Retirement to Glenlair

In 1865, at the age of 34, Maxwell decided to retire to his house, Glenlair, near Castle Douglas. He continued to work at electromagnetic theory, as well as running his estate and completing the extensions to the house that his father had planned. He was far too recognized an authority to be allowed to remain in isolation, and among the other opportunities offered him was the proposal in 1868 that he should become a candidate for the Principalship of St. Andrews. A fuller story of this incident has been given by Professor Domb in *Notes and Records of the Royal Society*,[23] but we may here note Maxwell's own assessment[31] of the danger of becoming Principal of a Scottish University. 'My proper line' he wrote 'is in working, not in governing, still less in reigning and letting others govern .....'

And while Aberdeen lost Maxwell, we may also record that St. Andrews not only lost him as Principal, but also lost James Joule, the discoverer of the conservation of energy, as its Professor of Natural Philosophy—for Lyon Playfair,[32] writing in 1890 to James Dewar, said that he had persuaded Joule to become a candidate for the St. Andrews Chair but 'He was on the point of securing this, but his personal slight deformity was an objection in the eyes of one of the electors, and St. Andrews lost the glory of having one of the greatest discoverers of our age'.

Maxwell continued to publish scientific papers during his retirement at Glenlair, and one of them—'On a Comparison of the Electric Units' in 1868—contains a note on the electromagnetic theory of light and shows how delicately he could deal with the errors of others. The mathematician Riemann and the physicist Weber had previously written papers based on action at a distance which suggested that electrical effects might be propagated with a velocity near that of light, but Maxwell[33] demonstrated that their theories were not physically viable, remarking 'From the assumptions of both these papers we may draw the conclusions, first, that action and reaction are not always equal and opposite, and second, that apparatus may be constructed to generate any amount of work from its resources ..... I think that these remarkable deductions from the latest developments of Weber and Neumann's theory can only be avoided by recognizing the action of a medium in electrical phenomena.'

Address to the British Association

In 1870 Maxwell was President of Section A of the British Association, and his Presidential Address[34] contains several famous passages:

> 'The human mind is seldom satisfied, and is certainly never exercising its highest functions, when it is doing the work of a calculating machine... There are, as I have said, some minds which can go on contemplating with satisfaction pure quantities presented to the eye by symbols, and to the mind in a form which none but mathematicians can conceive. There are others who feel more enjoyment in following geometrical forms, which they draw on paper, or build up in the empty space before them. Others, again, are not content unless they can project their whole physical energies into the scene which they conjure up. They learn at what a rate the planets rush through space, and they experience a delightful feeling of exhilaration. They calculate the forces with which the heavenly bodies pull at one another, and they feel their own muscles straining with the effort. To such men momentum, energy, mass are not mere abstract expressions of the results of scientific inquiry. They are words of power, which stir their souls like the memories of childhood. For the sake of persons of these different types, scientific truth should be presented in different forms, and should be regarded as equally scientific, whether it appears in the robust form and the vivid colouring of a physical illustration, or in the tenuity and paleness of a symbolical expression.'

Towards the end of the Address Maxwell made a prescient reference to the conflict of theories in physics:

> 'The most celebrated case of this kind is that of the corpuscular and the undulatory theories of light. Up to a certain point the phenomena of light are equally explained by both; beyond this point one of them fails... That theories apparently so fundamentally opposed should have

so large a field of truth common to both is a fact the philosophical importance of which we cannot appreciate until we have reached a scientific altitude from which the true relation between hypotheses so different can be seen.'

This passage reminds me of a debate[35] I had a few years ago with Sir Rudolf Peierls on the momentum associated with a beam of light: did it increase or decrease in an optically denser medium? Our views were in direct conflict, but each of us could point to a basic physical principle for support. When I pointed this out to him, Sir Rudolf wrote to me, 'We are in such a deep state of confusion that something good must come out of it.' What came out of the corpuscular-undulatory conflict was, of course, the photon—some fifty years after Maxwell had expressed his belief that the conflict would be resolved at a higher state of knowledge than obtained in his day.

Return to Cambridge: The Cavendish arises

Within a year Maxwell was back in harness and elaborating his views on the presentation of science in his Inaugural Lecture as the first Cavendish Professor in Cambridge. It was originally hoped that Kelvin might accept the Chair, but he felt too involved with his laboratory in Glasgow. Helmholtz was also approached, but the final choice fell on Maxwell, who was persuaded to move back to Cambridge from Glenlair. Not only had he to institute the new Department of Experimental Physics, but he had also to design the Cavendish Laboratory. Typically, his sense of mischief was manifest on the occasion of his Inaugural Lecture. The lecture itself was a magnificent exposition of the place of experimental physics in a university, and of the principles of teaching and research. In teaching he proposed to mobilize all the methods which he had outlined in his British Association Address while in research he spoke[36] about the need for precise measurements:

'This characteristic of modern experiments—that they consist principally of measurements—is prominent, that the opinion seems to have got abroad that in a few years all the great physical constants will have been approximately estimated, and that the only occupation which will then be left to men of science will be to carry on these measurements to another place of decimals.

'If this is really the state of things to which we are approaching, our Laboratory may perhaps become celebrated as a place of conscientious labour and consummate skill, but it will be out of place in the University, and ought rather to be classed with the other great workshops of our country, where equal ability is directed to more useful ends.

'But we have no right to think thus of the unsearchable riches of creation, or of the untried fertility of these fresh minds into which these riches will continue to be poured. It may possibly be true that, in some of those fields of discovery which lie open to such rough observations as can be made without artificial methods, the great explorers of former times have appropriated most of what is valuable, and that the gleanings which remain are sought after, rather for their abstruseness, than for their intrinsic worth. But the history of science shews that even during that phase of her progress in which she devotes herself to improving the accuracy of the numerical measurement of quantities with which she has long been familiar, she is preparing the materials for the subjugation of new regions, which would have remained unknown if she had been contented with the rough methods of her early pioneers. I might bring forward instances gathered from every branch of science, shewing how the labour of careful measurement has been rewarded by the discovery of new fields of research, and by the development of new scientific ideas.'

The Cavendish Laboratory, Free School Lane, Cambridge (by permission of the Director of the Cavendish Laboratory).

Some of Maxwell's shafts were directed against the attitude that we have already seen typified by Todhunter. Indeed, in an essay on 'The Conflict of Studies,' Todhunter opposed the introduction of experimental illustrations into lectures, on the grounds that an experiment which is not intended to bring out a new fact is useless, and as for experiments of demonstration, 'It may be said that the fact makes a stronger impression on the boy through the medium of his sight, that he believes it the more confidently. I say that this ought not to be the case. If he does not believe the statements of his tutor—probably a clergyman of mature knowledge, recognized ability, and blameless character—his suspicion is irrational and manifests a want of power of appreciating evidence. A want vital to his success in that branch of science which he is supposed to be cultivating.'[17]

Todhunter, though, was probably not present at Maxwell's Inaugural Lecture for, whether by accident or design, it was poorly advertised. When Maxwell's first lecture to his students was announced, the senior members of the university therefore assumed that this would be his Inaugural Lecture, and they turned up in force. His students therefore had to stay in the back rows watching with some delight as Maxwell gave his first elementary lecture on heat while such men as the Vice Chancellor, Cayley, and Stokes sat in the front row while Maxwell gravely expounded to them, though with a twinkle in his eye, the difference between the centigrade and Fahrenheit scales.

Maxwell's views on the ways in which information could be conveyed orally were whimsically expressed in a disapproving letter[37] of 1873 to W. G. Adams on the proposal to found the Physical Society:

> 'For the evolution of science by societies the main requisite is the perfect freedom of communication between each member and any one of the others who may act as a reagent. The gaseous condition is exemplified in the soiree, where the members rush about confusedly... The opposite condition, the crystalline, is shown in the lecture, where the members sit in rows, while science flows in an uninterrupted stream from a source which we take as the origin. This is radiation of science. Conduction takes place along the series of members seated round a dinner table, and fixed there for several hours with flowers in the middle to prevent any cross currents. The condition most favourable to life is an intermediate plastic or colloidal condition...'

1873 saw the publication of the great work which had occupied much of Maxwell's time for the past several years, the 'Treatise on Electricity and Magnetism,' in which he set forth his electromagnetic theory in its final form. Besides obtaining more elegantly all his earlier results, it now included a new one of great significance in its Article No. 792. Ever since Newton had propounded the corpuscular theory of light, it seemed that the light corpuscles ought to carry momentum, and therefore that light should exert an appreciable pressure on any surface on to which it fell. In the preceding two centuries it had been the object of experimenters to demonstrate the existence of this pressure, but they had all failed. It was indeed doubted whether such pressure existed (and Thomas Young[38] had erroneously used the apparent lack of radiation pressure as an argument in favour of the wave theory), but Maxwell was now, for the first time, able to calculate its magnitude. He had formed a mental picture of what an electromagnetic wave might be like, as consisting of transverse electric and magnetic vibrations at right angles to the direction of propagation, and he was able to calculate the current that would be induced in a mirror surface by the fluctuating magnetic component. The induced current would develop a fluctuating magnetic field which would interact with that in the wave itself, and the interaction would take the form of a force of repulsion. Maxwell calculated the magnitude of the force and found that even in full daylight it would amount to only about 1 lb weight on a half mile square. The reason, therefore, that the force had so far remained undetected was that it was too small for any previous experimenters to have observed; and it was not until 26 years later, in 1899, that the Russian physicist P. Lebedew[39] succeeded in refining experimental technique to the stage where Maxwell's prediction could be confirmed. Although the pressure is so very small in ordinary circumstances, it is otherwise in the interiors of stars, where there are enormous quantities of radiation, and Eddington was able to show that it is a major factor in the structure of stars, where the outer layers are prevented from collapsing inwards because of the radiation exerted upon them from the inside. It is when the radiation begins to weaken as a result of the nuclear furnace failing, that the star collapses and becomes a nova or super-nova. The existence of radiation pressure of light also had a further significance, in that Einstein[40] was able to use it in a 'thought-experiment' by which he could persuade his contemporaries of the validity of the famous equation $E = mc^2$.

James Clerk Maxwell, F.R.S., 1831-1879.

The Last Years

Maxwell's original work was now largely over—what he might have done, had he not carried the burden of the new Chair, we do not know. He saw that his first aim must be to set the Cavendish Laboratory going on the right lines; and, in return for the sacrifice of his own work, the world gained its most famous laboratory. Again, Maxwell did not shrink from duty in the writing of textbooks, even elementary ones. They included, besides the great 'Treatise', elementary textbooks of authority and clarity on 'Matter and Motion', and 'The Theory of Heat', in which he described the famous Maxwell demon.

Even so, there were some very substantial papers in his later years. The Royal Society of Edinburgh awarded him its Keith Prize in 1873 for his work on reciprocal diagrams applied to engineering structures,[41] and its extension to the mechanics of continuous media. He also returned to the kinetic theory of gases, and in two of his most powerful papers (1879) extended Boltzmann's

Principle of the Equipartition of Energy from molecules in gases to those in liquids and solids,[42] and he worked out the theory of the radiometer newly discovered by Crookes.[43]

For a time it was believed that the radiometer worked by the electromagnetic radiation pressure that Maxwell had predicted, but the effect was much too large and it turned out to have an entirely different explanation. Maxwell was called in 1876 to demonstrate the radiometer to Queen Victoria and he described the experience[44] in characteristic style: 'I was sent for to London, to be ready to explain to the Queen why Otto von Guericke devoted himself to the discovery of nothing, and to show her the two hemispheres in which he kept it, and the pictures of the 16 horses who could not separate the hemispheres, and how after 200 years W. Crookes has come much nearer to nothing and has sealed it off in a glass globe for public inspection. Her Majesty, however, let us off very easily and did not make much ado about nothing, as she had much heavy work cut out for her all the rest of the day...'

There was evidence, too, in his last years that Maxwell's interest in scientific instruments continued. When the collection of scientific instruments went on exhibition at South Kensington in 1876, he wrote the introductory chapters on the principles of instrument design for the accompanying handbook.[45] It is notable for the first statement of the principles of kinematic design and of flexure movements.

Maxwell's last years were saddened by the continuous illness of his wife, whom he nursed himself. At one time he did not go to bed for three weeks but conducted his lectures as usual. From 1877 onwards he himself showed increasing signs of ill health, and in October 1879 he was told at Glenlair that he had not a month to live. He returned to Cambridge and died on 5th November, retaining his unselfishness and composure to the end. He was a victim of the same disease that had killed his mother at the same untimely age of 48, cancer of the stomach. He was buried in Parton Churchyard a few miles from Glenlair, beside Loch Ken.

A memorial plaque and a stained glass window were erected in the Maxwell Kirk near Glenlair; when this Kirk fell into disuse they were both transferred to the Church at Corsock. One day in 1964, I happened to be at lunch with the late Lord Reith, and on learning of my interest in Maxwell, he told me that he had come across the memorial plaque to Maxwell in Corsock Church, which led me to recall a story which he had recounted in his autobiography[46] concerning a certificate which hung on the wall of his office throughout the time that he was head, first of the British Broadcasting Company and later of the British Broadcasting Corporation. With his known reputation for piety, he nevertheless appreciated the general belief held by his staff that it was a Sunday School Certificate won by himself. In fact, it was a Merit Certificate for a member of the Natural Philosophy class at Aberdeen in 1860, signed by James Clerk Maxwell. The winner of the Certificate was George Reith, later Moderator of the Church of Scotland, and father of Lord Reith, who saw it as his heaven-sent destiny to put to work for the public good the electromagnetic waves which his father's professor had predicted. Sensing my interest in the Certificate he asked me whether I would like to have it. When I said that I would, especially since we had no Maxwell relics in Aberdeen, he replied 'You shall have it. As a matter of fact I gave it to Sir Noel Ashbridge, but I'll make him give it back' and before the afternoon was over a subdued Sir Noel telephoned me to tell me that Lord Reith had instructed him to hand it over.

Incidentally, Lord Reith also sent me a copy of a note written by his father in 1914 on his days at Aberdeen Grammar School and Marischal College, commenting on his professors: 'But much more notable there was Clerk Maxwell, a rare scholar and scientist as the world came to know afterwards; a noble-souled Christian gentleman with a refined delicacy of character that bound his class to him in a devotion which his remarkably meagre qualities as a teacher could not undo.' To which Lord Reith added a manuscript note 'I'm sure my father's use of "gentleman" was deliberate.'

Glenlair, Maxwell's home, stands in ruins. Taking its name from the Lair Burn, beside which it was built by Maxwell's father, it was extended after his death by Maxwell himself from plans which his father had drawn up: and it was at Glenlair that Maxwell conceived the electromagnetic theory. When he first worked out the velocity of his electromagnetic waves, and found that it was very near the velocity of light, he wrote[47] to Faraday on 19th October 1861: 'This coincidence is not merely numerical. I worked out the formulae in the country before seeing Weber's number, which is in millimetres, and I think we have now strong reason to believe, whether my theory is a fact or not, that the luminiferous and the electromagnetic medium are one.' Later on 10th December of the same year he wrote to Kelvin,[48] 'I made out the equations in the country before I had any suspicion of the measures between the two values of the velocity of propagation of magnetic effects and that of light.'

Unfortunately, the Maxwells had no children, and the estate passed into other hands. A disastrous fire in the 1920s turned the house into a ruin which it would be impracticable to rebuild. With the world-wide recognition of Maxwell's genius, the most that we might hope is that one day the ruin might be cleaned up and become a source of pilgrimage just as the English Abbeys have become.

Assessment of the Physicist and the Man

When Maxwell died the electromagnetic theory was by no means entirely accepted. Even in 1884, Kelvin was saying[49] in his famous Baltimore Lectures 'The so-called "electromagnetic theory of light" has not helped us much hitherto'. Even the famous displacement current was the subject of some doubt—for how could a vacuum carry an electric current, even for an instant?—and the Berlin Academy (1879) had offered a prize for its experimental verification. Maxwell himself in his Note of 1868 had said that 'The current produced in discharging a condenser is a complete circuit, and might be traced within the dielectric itself by a galvanometer properly constructed. I am not aware that this has been done, so that this part of the theory, although apparently a natural consequence of the former, has not been verified by direct experiment. The experiment would certainly be a very delicate and difficult one'. Helmholtz suggested to Heinrich Hertz that he should try to win the prize by detecting the displacement current, but Hertz considered the experiment too difficult.[50] A few years later, though, he realized that it might be possible to generate electromagnetic oscillations in free space, and so he set about trying to make these oscillations and thus confirm the electromagnetic theory directly. In the year 1886 he performed the brilliant series of experiments which established his own fame and confirmed the fame that up to that time had only reservedly been given to Maxwell's theory.[51] From Hertz's experiments Marconi realized the possibility of radio communication, which had indeed almost certainly been anticipated by David Hughes in 1879, but its nature had not at that time been realized by Hughes or anybody else,[52] and Marconi's work led on to all the applications of radio waves that we know today, including radar and television.

As for Maxwell's character, his delightfully whimsical humour and deep clarity of thought are evident from his writings. Shy, modest, gentle, kind, and noble, he aroused no animosities and inspired no jealousies. A devoted husband, an enthusiastic if not lucid teacher, sensitive to history and unafraid of metaphysics, he was deeply religious. The memorial window at Corsock carries the Greek version of the words[53] he quoted on his deathbed: 'Every good and perfect gift comes from above' adding 'Do you know that that is a hexameter in Greek?'

One of his last comments[54] was 'I have looked into most philosophical systems, and I have seen that none will work without a God.' Deeply Christian, and an Elder of the Kirk, he was the main contributor to the endowment of Corsock Church and the building of its manse. But his religion had a deeply personal aspect: 'I think that the results which each man arrives at in his attempts to harmonize his science with his Christianity are not to be regarded as having any significance except to the man himself, and to him only for a time, and should not receive the stamp of a society.'[55] Although he saw the case for authoritarianism in religion, he was against it: 'People get tired of being able to do as they like, and having to choose their own steps, and so they put themselves under holy men, who, no doubt, are really wiser than themselves. But it is not only wrong, but impossible, to transfer either will or responsibility to another.'[56]

In his Inaugural Lecture at Cambridge, Maxwell had repeated his interest in the History of Science, since it involved the appreciation of human character as well as of scientific thought: 'The men whose names are found in the History of Science are not mere hypothetical constituents of a crowd, to be reasoned upon only in masses. We recognize them as men like ourselves, and their reactions and thoughts, being more free from the influence of passion, and recorded more accurately than those of other men, are all the better materials for the study of the calmer parts of human nature.'

Few of us would presume to recognize in Maxwell a man 'like ourselves': he was so much superior as a physicist, and exemplary as a human being. Let us leave that privilege to men like Planck and Einstein. At the 1931 celebrations of the Centenary of Maxwell's birth, Planck said[57] 'It was his task to build and complete the classical theory, and in doing so he achieved greatness unequalled. His name stands magnificently over the portal of classical physics.'

Einstein, in his autobiographical notes, recorded that[58] 'The most fascinating subject at the time that I was a student was Maxwell's theory. What made this theory appear revolutionary was the transition from forces at a distance to fields as fundamental variables. The incorporation of optics into the theory of electromagnetism was like a revolution....'and he opened his famous 1905 paper on Relativity by pointing out how he had been struck by the asymmetrical explanations of the phenomenon of inducing a current in a circuit with a magnet, depending on whether one regards the magnet as moving and the circuit fixed or vice versa. And later in the same notes, Einstein simply said 'The special theory of relativity owes its origin to Maxwell's equations of the electromagnetic field'. And in the 1931 celebrations Einstein returned to the point that had so impressed him as a student: 'Since Maxwell's time, Physical Reality has been thought of as represented by continuous fields . . . and not capable of any mechanical interpretation. This change in the conception of Reality is the most profound and the most fruitful that physics has experienced since the time of Newton.[59]

So the man whose memory we honour tonight stood at the pinnacle of classical physics and pointed the way for the great developments that have since occurred. Besides his conception of the Electromagnetic Theory, he illuminated every branch of physics which attracted his attention: kinetic theory, thermodynamics, colour vision, cybernetics, optical and mechanical instrument design, and structural engineering. With him, experiment went hand in hand with theory; and he was happy in design, whether it be of an instrument, an experiment, or the Cavendish Laboratory. Acutely sensitive to the history of his subject, he stands in it as complete, and as perfect, a physicist as has ever lived.

References

1 Campbell, L., and Garnett, W., 'The Life of James Clerk Maxwell', 1st ed. (Macmillan, London, 1882) (Cited in subsequent references as 'Life').
2 Niven, W. D. (Ed.) 'The Scientific Papers of James Clerk Maxwell', Volumes I and II, (Cambridge University Press, 1890; reprinted in facsimile by Dover, New York, 1965', (Cited in subsequent references as 'Papers'). I, p. 1.
3 Reid, T., 'An Inquiry into the Human Mind on the Principles of Common Sense' (1764).
4 'Life', p. 116.
5 'Papers', II, p. 233.
6 'Papers', I, p. 155.
7 'Life', p. 235.
8 Foucault, L. and Fizeau, H. L., *Comptes Rendus*, **30**, pp. 551–662, 1850.
9 Faraday, M. 'Experimental Researches in Electricity', Vol. III, p. 447 (Quaritch, London, 1855).
10 'Life', p. 268.

11 'Papers', I, p. 136.
12 'Life', p. 263.
13 Jones, R. V., *Notes and Records of the Royal Society*, **28**, pp. 57–81, 1973.
14 Lamb, H., in 'James Clerk Maxwell, A Commemoration Volume', pp. 144, 145 (Cambridge University Press, 1931) (Cited in subsequent references as 'Commemoration Volume').
15 'Life', p. 297.
16 'Life', pp. 295, 301, 303.
17 Schuster, A., 'The Progress of Physics', pp. 25–26 (Cambridge University Press, 1911).
18 'Papers', I, p. 427.
19 'Papers', I, p. 277.
20 'Life,' p. 303.
21 'Papers', II, p. 377.
22 Walker, R., *Aberdeen University Review*, **3**, p. 193, 1916.
23 Domb, C., *Notes and Records of the Royal Society*, **35**, pp. 67–103, 1980.
24 'Papers', I, p. 491.
25 Page, L. and Adams, N. I. A., 'Principles of Electricity', 4th Edn., p. 471 (Van Nostrand, New York, 1969).
26 'Papers', I, p. 490.
27 'Papers', I, p. 529.
28 'Papers', II, pp. 1–25.
29 Lord Rayleigh, *Nature*, **28**, pp. 26–27, 1890.
30 'Papers', II, pp. 105–120.
31 Clerk Maxwell, J., Letter of 30th October 1868 to Kelvin. (Courtesy of Mr J. T. Lloyd, Keeper of the Kelvin Museum, Glasgow.)
32 Reid, W., 'Memoirs and Correspondence of Lyon Playfair', p. 74 (Cassell, London, 1899).
33 'Papers', II, pp. 137–38.
34 'Papers', II, pp. 215–29.
35 R.V.J. Correspondence.
36 'Papers', II, pp. 241–55.
37 'Life', p. 384.
38 Young, T., *Phil. Trans. Roy. Soc.*, **92**, p. 12, 1802.
39 Lebedew, P., *Archives des Sciences Physiques et Naturelles*, **8**, p. 184, 1899.
40 Einstein, A., *Annalen der Physik, Leipzig*, **20**, p. 627. 1906.
41 'Papers', II, pp. 161–207.
42 'Papers', II, pp. 713–41.
43 'Papers' II, pp. 681–712.
44 Everitt, C. W. F., 'James Clerk Maxwell', p. 157 (Scribner's, New York, 1975).
45 Clerk Maxwell, J., 'South Kensington Museum Handbook to the Special Loan Collection', pp. 1–21 (Chapman and Hall, London, 1876).
46 Reith, J. C. W., 'Into the Wind' (Hodder & Stoughton, London, 1949).
47 Campbell, L., and Garnett, W., 'The Life of James Clerk Maxwell', 2nd ed., p. 244 (Macmillan, London, 1884).
48 Larmor, J., *Proc. Camb. Phil. Soc.*, **32**, pp. 695–750, 1936.
49 Lord Kelvin, 'Baltimore Lectures', Preface, p. vii. (Cambridge University Press, 1904).
50 Whittaker, E. T., 'A History of the Theories of Aether and Electricity: The Classical Theories', 2nd Edn., p. 322 (Nelson, London, 1951).
51 Hertz, H., *Ann. d. Phys.*, **31**, p. 421, 1887.
52 Burniston Brown, G., *Notes and Records of the Royal Society*, **34**, pp. 227–39, 1980.
53 'Life', p. 409.
54 'Life', p. 426.
55 'Life', p. 405.
56 'Life', pp. 306–7.
57 Planck, M., 'Commemoration Volume', p. 65.
58 Einstein, A., in 'Albert Einstein, Philosopher Scientist', Schilpp, P. A. (Ed.), Vol. 1, p. 33 (Harper, New York, 1959).
59 Einstein, A., Commemoration Volume, p. 71.

The Author

Professor R. V. Jones was educated at Wadham College and the Clarendon Laboratory, Oxford, where he took his D.Phil. degree in Infra-red Spectroscopy in 1934. He was subsequently Skynner Research Student in Astronomy at Balliol College, and worked on the detection of aircraft by infra-red radiation in the years before World War II. During the war he was Head of Scientific Intelligence, and was Director of Intelligence on the Air Staff and later Director of Scientific Intelligence in the Ministry of Defence. His experiences are described in his book 'Most Secret War', published in 1978. He was appointed to the Chair of Natural Philosophy at Aberdeen University in 1946.

Many governmental and other bodies have called upon Professor Jones's knowledge and experience during the past thirty-five years. His appointments have included the Chairmanship of the Safety in Mines Research (Advisory) Board, of the Research Advisory Council of the Transport Commission, of the Paul Instrument Fund Committee, of the Infra-red Committee in the Ministry of Aviation, of the Electronics Research Council in the Ministry of Technology, of the Air Defence Working Party of the Ministry of Defence, and of the British National Committee for the History of Science, Medicine and Technology. He has been Vice-President of the Royal Society and is Editor of the Society's *Notes and Records*. He is an Honorary Fellow of Wadham and Balliol Colleges, and of the College of Preceptors, and has Honorary Doctorates from the Universities of Strathclyde, York, Bristol, Surrey and Kent and from the Open University.

His research interests include particularly the measurement of small displacements, and fundamental phenomena in optics such as radiation pressure and the aberration of light.

Contributors to this Issue

Philip Williams was educated at the then University College of Nottingham and served in the Royal Navy from 1943 to 1946. He has been working on radar and kindred topics since 1943, and has been with Decca Radar since 1952. In 1964 Mr Williams carried out a collaborative research between University College London and the former Decca Radar Company which resulted in the introduction of the 'Clearscan Processor' for surveillance radars and gained him an M.Phil. degree. Mr Williams is now a Research Fellow at the University of Surrey, concerned with various projects in the infra-red and microwave spectrum. In the radar field his interest in sea ice detection has taken him to the Baltic, Labrador Straits and the Beaufort Sea to carry out both airborne and ship-borne experimental work. Arising out of this work, he was awarded the Brabazon Premium for a paper published in the Journal in 1979. Mr Williams is the author of numerous other papers and holds 12 patents.

Raymond Steele was an indentured apprenticed radio engineer before going to Durham University, where he attained a bachelor's degree in electrical engineering in 1959. After research and development posts with E. K. Cole, Cossor Radar and Electronics, and the Marconi Company, he joined the lecturing staff at the Royal Naval College, Greenwich. Here he lectured on Telecommunications to the NATO and the External London University degree courses. As a Senior Lecturer in the Electronic and Electrical Engineering Department of Loughborough University of Technology, he directed a research group in digital encoding of speech and television signals. In 1975 he received his Ph.D. Dr. Steele was a consultant to the Acoustics Research Department at Bell Laboratories in the summers of 1975, 1977 and 1978 and in 1979 he joined the Company's Communications Methods Research Department. He is the author of the book, 'Delta Modulation Systems', has written over 50 papers relating to digital encoding and signal processing, and was awarded the Institution's Marconi Premium for 1978.

Wai Choong (Lawrence) Wong received the B.Sc. degree in Electronic and Electrical Engineering and the Ph.D. degree in Electronic Engineering both from Loughborough University of Technology, Loughborough, England, in 1976 and 1980 respectively. Since April 1980 he has been a Member of Technical Staff in the Communications Methods Research Department of Bell Laboratories at Crawford Hill, Holmdel, New Jersey, U.S.A. His current interest is in speech signal processing particularly applied to mobile radio environment.

Tony Henk (Member 1973) joined the BBC Television Service in 1954, working on operations and maintenance at Lime Grove Studios. In 1956 he moved to Midlands Independent Television where he was responsible for the maintenance team at ITV Midlands Studio centre and for design and development of sound and vision equipment, including original work on inlay and overlay techniques. Eleven years later he went over to research work with the Plessey Company at the Roke Manor laboratory, carrying out v.h.f. and u.h.f. mobile radio propagation measurements and modelling during preliminary phases of Project Mallard. In 1956 Mr Henk joined the then British Radio Corporation at Bradford as Chief Development Engineer, leading the team responsible for the 8000 and 8500 colour receivers which pioneered the use in consumer television in the UK of synchronous vision demodulation and a single-transistor line output stage. During this period he qualified for membership of the Institution by means of a thesis on the 8000 Receiver.

After some six years in manufacturing industry Mr Henk went to Burmah Engineering in 1975 as part of a team of telecommunications specialists providing consultancy services to oil and gas companies planning offshore platforms. A year later he took up his present appointment with Brown & Root (UK) as Senior Specialist Telecommunications Engineer responsible for planning and design of telecommunications systems for the company's clients in the on-shore and off-shore oil and gas industries throughout the world.

Professor Richard Waldron (Fellow 1961, Member 1957) has been Head of the School of Mathematics at Ulster Polytechnic with the title of Director of Studies since 1973; he was awarded the title of Professor two years ago. Before going to Northern Ireland he held a three-year Principal Research Fellowship at the Post Office Research Centre, Martlesham, where he was mainly concerned with trunk waveguide communications.

Professor Waldron graduated from the University of Cambridge in 1951 and joined the Baddow Research Laboratories of the Marconi Company, where he worked mainly on microwave theory. During this period he contributed numerous papers on waveguide subjects to the Institution's Journal for several of which he received premiums. In 1968 he was awarded the degree of Sc.D. from Cambridge University on the basis of this and other published work. From 1968 to 1971 Professor Waldron held a research appointment at the Lincoln Laboratories of the Massachusetts Institute of Technology, working on surface-acoustic-wave theory. His publications also include the expression of some unorthodox views on relativity theory, on which he has spoken at seminars in several universities and elsewhere. In addition to his many papers, he has published books on ferrites and on the wave and ballistic theories of light.

During the past ten years, Professor Waldron has been active in Local Section affairs; he has held office as both Programme Secretary and Honorary Secretary of the East Anglian Section, and he is currently Chairman of the Northern Ireland Section.

UDC 621.396.963.325: 621.391.81
Indexing Terms: Radar, marine, Radar interference, clutter, Diversity.

Results from an experimental dual-band search radar

P. D. L. WILLIAMS, M.Phil., C.Eng., F.I.E.E., S.M.I.E.E.E.*

SUMMARY

The performance of surface surveillance radars is complicated by temporal changes of propagation loss, sea and rain clutter and target behaviour. This paper presents short and medium-term statistics using a real time pulse-to-pulse digital analyser operated in conjunction with high dynamic range logarithmic receivers fed from an experimental dual antenna coastal radar.

Previous performance predictions[18] made for equipments operating in true diversity, themselves supported by extensive trials[24] over half minute intervals of observation, are now augmented with true pulse-by-pulse measurements on targets obtained in each common beam dwell time of a true composite dual channel experimental radar.

Short-term measurements on sea and rain clutter show that predictions based on average or mean values of sea and rain clutter are a poor guide to true short term variations, but become more accurate over longer periods as found by others.[25]

Illustrations from p.p.i.s are given of both targets and clutter from the experimental radar, showing the effects of optimum and non-optimum video processing of the output from the two channels and results from a captive target balloon at selected elevations are included.

* *Visiting Fellow, Department of Electronic and Electrical Engineering, University of Surrey, Guildford GU2 5XH.*

1 Introduction

The design of a radar system requires many parameters to be examined. High on the list is the choice of carrier frequency and usually the transmitted waveform and its spectrum is considered separately. This paper discusses the advantage to be gained by a split spectrum transmitted waveform where the power generation is carried out by virtually two independent power oscillators operating in time synchronization but with a very large carrier frequency separation.

For a number of reasons the choice of a single operating frequency is often far from clear and the use of more than one carrier frequency has many advantages. True frequency diversity as opposed to frequency agility allows simultaneous transmission of the two or more carriers and in some ways eases the signal combination problem. It is the purpose of this paper to report on results obtained from an experimental twin-carrier system operating with a 3-to-1 carrier separation at wavelengths in the 10 cm and 3 cm bands.

In addition to the use of two wavelengths the two aerial beams (which are aligned in azimuth) had their vertical directions deliberately offset so as to maximize detection of aircraft whilst maintaining the best surface cover for shipping and navigation purposes for a sea-going system. The higher frequency carrier has the nose of its vertical aerial beam set horizontally and the lower frequency carrier has its nose set at $+20^{\circ}$ for these experiments; this arrangement still allowed a maritime version to roll and pitch $\pm 7^{\circ}$ without loss of surface cover, which using the high resolution beamwidth allows good mapping capabilities for coastal navigation.

2 Background and Prior Work

In this specific radar area the literature falls into two groups. The first group covers the radar reflecting properties of targets such as aircraft, shipping, etc., and all the various forms of clutter. This classification is, of course, confused by the fact that one system designer's 'target' may be another's 'clutter'. Rain and other forms of precipitation, land and coastal features, the sea surface and any ice covering, all have a dual identity. The second group of literature is that devoted to the theoretical and practical aspects of multi-carrier or spread spectrum radars. Examples from the first group are given in the first references, which include some modern reviews and reprints of famous papers, and also a colloquium[10] held in London on 'The Sea Surface,' when those people regarding it as a target met those who regarded it as unwanted clutter.

The second group in the literature specifically addresses either true diversity operation of radar systems or measurements on targets and clutter made from colocated radar equipment at the same time. The most comprehensive work in this second category must be the '4FR' trials carried out by NRL Washington, and Valenzuela[11–13] in his review of the air–ocean interface

0033–7722/81/110541+12 $1.50/0

provides a recent and convenient entry to that published work by Daly[14] *et al.*

Further examples of multi-frequency radar measurements of the sea surface have been reported by Sittrop,[15] Dyer and Currie,[16] Lewis and Olin,[17] and Williams,[18] whilst theoretical aspects of target detection in diversity systems are given by Vannicola[19] and Poelman;[20] Barton[21] has edited a modern collection of papers with comments on availability of some publications in this field,[22] Williams[23] has reported on target trials using 10 cm and 3 cm radars operating from a cliff top overlooking the sea some 45 metres above sea level where that particular paper was concerned with surface target behaviour under calm dry conditions without sea or rain clutter being present.

3 Objectives

This paper reports on fresh measurements carried out with a twin-carrier, pulsed, coastal radar. In addition to observations on visual displays, pulse return statistics on a selection of radar targets and clutter are also presented. The first period of observations was accompanied by high winds and rain storms to provide an interesting range of clutter backgrounds. Later observations were carried out in calmer weather needed for the tethered balloon runs.

4 Experimental Equipment

The two parts of a composite aerial consisting of two horizontally polarized, slotted waveguide aerials rotated on a common turning gear were coupled to two transceivers. The separate resultant video outputs are available for use either on separate p.p.i. displays or for composite presentation in much the same way as has been previously reported for use on a special 'split image' display.[18–23] Additionally the amplitude of all pulses could be staticized and recorded during the trials.

The essential parameters, together with the expected range performance on a non-frequency sensitive target of radar cross-section (r.c.s.) equal to 1 m² (σ_{50}), are given in Table 1.

5 Performance Prediction

5.1 Performance Prediction in Free Space

If the target has a median value of 1 m² at both wavelengths and the equipment operates on the medium pulse length of 0·25 μs, then the expected maximum detection range is 12·6 km for the 3·2 cm carrier and 13·5 km for the 10 cm carrier. The effect of rain clutter may also be estimated from the size of radar resolution cell (at a given range) and appropriate backscatter coefficient conveniently taken from Nathanson's tabulations.[3] The absolute values of rain backscatter are dependent on many natural parameters such as actual rainfall in the radar beam as opposed to that measured on the ground, drop size distribution and temperature, but an estimate is sufficient to indicate the likely relative performance on each of the two carriers.

Table 1.
Equipment parameters

Parameter	Carrier Wavelength	
	3·2 cm	10 cm
TRANSMITTER		
Peak power	25 kW	30 kW
Pulse length choice	0·05 μs 0.25 μs 1 μs	0·05 μs 0·25 μs 1 μs
Corresponding p.r.f.	3300 Hz 1650 Hz 825 Hz	3300 Hz 1650 Hz 825 Hz
Instrumental limitation of p.r.f. (Trials only)	850 Hz	850 Hz
Wavelength	3·2 cm	10 cm
RECEIVER		
(Type: Superhet with logarithmic i.f. amplifier)	—	—
Noise factor (overall)	10 dB	10 dB
Bandwidth: long and medium pulse	5 MHz	5 MHz
short pulse	15 MHz	15MHz
AERIALS		
Trials height a.s.l.	20 m	19·5 m
Rotation rate	20 rev/min	— common
Physical aperture:—	9 ft (2·8 m)	12 ft (3·75 m)
Effective horizontal	2·9 m	3·76 m
Effective vertical	0·29 m	0·5 m
Polarization	H	H
Gain	33 dB	27 dB
Horizontal beamwidth	0·85°	2°
Vertical beamwidth	15°	32°
Estimated Losses	5 dB	5 dB
Beam dwell time on a point target	7·1 ms	16·7 ms
For a p.r.s. of 850 the pulses per beamwidth (N) =	11·6	27·5
Required S/N for P_d = 50%	5 dB	2·5 dB
Calculated free space range on σ_{50} = 1 m²	12·6 km	13·5 km
Transition range for target 1 m a.s.l	7·5 km	2·4 km

5.2 Performance Prediction in Rain

Table 2 lists these parameters and gives the required estimates of competing rain clutter, assuming a local storm and no excess attenuation between the radar and the rainstorm, particularly on the lower wavelength

Table 2
Parameters controlling rain clutter assessments

Parameter	Carrier Wavelength	
	3·2 cm	10 cm
Azimuth beamwidth	2°	0·85°
Elevation beamwidth	15°	32°
Range resolution	0·25 μs (38 m)	0·25 μs (38 m)
Range for rain clutter comparison	10 km	10 km
Volume of radar resolution cell	$14·5 \times 10^6$ m³	73×10^6 m³
Rainfall rate assumed	4 mm/hr	4 mm/hr
σ_0 for above rate	−62 dB m²/m³	−83 dB m²/m³
Effective rain clutter r.c.s.	9·2 m²	0·4 m²

carrier. If in fact the rain is continuous for 10 km an extra propagation loss of 6·9 dB will be suffered for targets and clutter at the higher carrier frequency.

5.3 Performance Prediction in Sea Clutter

Sea clutter comparisons are very difficult owing to the greater spike content from the higher resolution 3·2 cm carrier and, of course, for moderate common aerial heights above the sea, the transition range for the lower frequency carrier ensures a much shorter range cut-off. Thus the lower frequency carrier is preferable for high target detection in both sea and rain clutter as well as in a thermal noise background, but for low lying surface targets and modest aerial heights the short transition ranges[18] become a source of serious loss at all but very modest distances from the radar.

5.4 Environmental Diagram

The relative merits and demerits of each carrier irrespective of aerial gain on the water are shown in Figs. 1 and 2, prepared by calculating the values of mean rain and sea clutter and assuming a signal-to-clutter ratio of 10 dB. The transition ranges given in Table 1 are also shown and mark the apparent falling off of a surface target of radar cross-section $= 1\ m^2$ beyond these transition ranges to an R^{-4} law as reported by Williams.[18]

Sea clutter equivalent r.c.s. mean values are also shown and it will be seen that the lower values of σ_0 at S-band have outweighed the larger aerial beamwidth to still produce less sea clutter compared to a $1\ m^2$. We have assumed a complex target where the median values are the same at both wavelengths.[18]

The S-band set shows that a $1\ m^2$ target 1 m above sea level will conflict with the rain clutter at 4 mm/hr at 3·5 km. Although at X-band the performance against the same target stays inside the transition range to 7·5 km

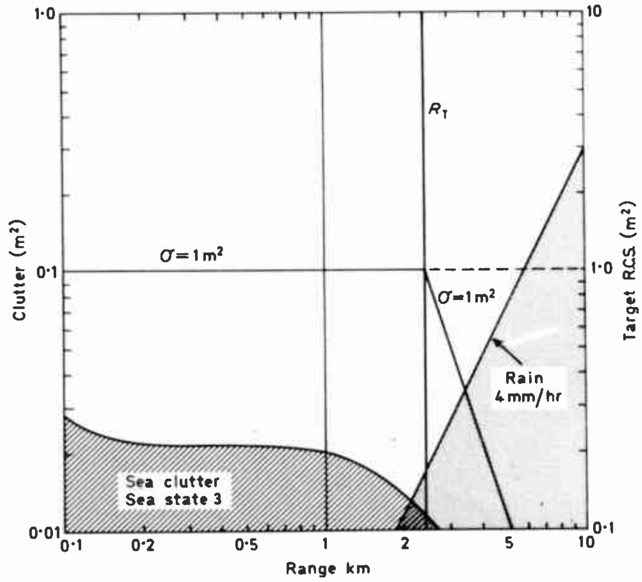


Fig. 1. Environmental diagram for 10 cm element.

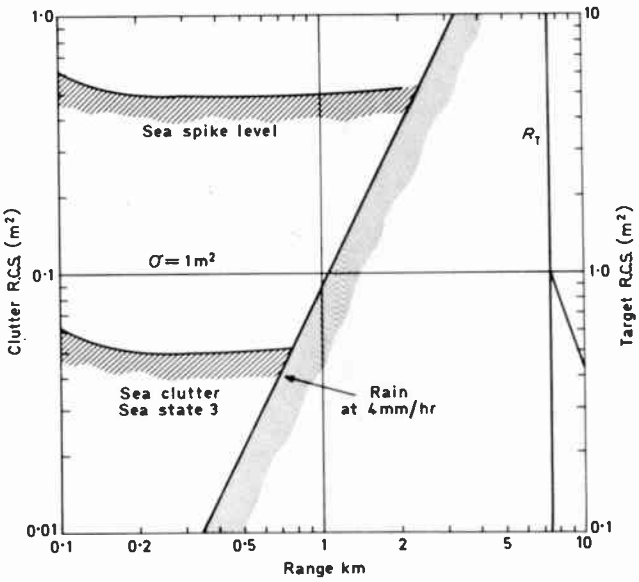


Fig. 2. Environmental diagram for 3·2 cm element.

and is likely to be seen against noise out to double that range, it can be obscured by rain beyond only 1 km and has not such a clear margin above even the mean sea clutter. The high value 'sea spikes' reported by Olin and Lewis,[17] Valenzuela,[11] and Williams[18] are likely to give trouble in the form of excessive false alarms at all ranges to the rain model for 4 mm/hr.

If a practical target having 4 metres freeboard is considered, then at 5 km range there will be some 4 lobe maxima effectively illuminating the target at X-band and only one at S-band, so that it is more likely that part of the X-band lobe pattern will strike a strong echo centre on the target than the single maxima from the S-band aerial. To model this needs a very detailed knowledge of the spatial distribution of echo centres for all azimuths and vertical aspects of any target, not a practical task upon which to embark. The transition range concept is far more simple to use where targets do extend over several vertical lobes, unlike aircraft at 100 km or more which fall into the lobe maxima and minima from the interferometer set up by a surface aerial and its image beneath.

With this and previous theoretical background,[4] after a brief description of the pulse analyser, real results will be considered.

6 Additional Instrumentation

Various methods of gathering amplitude data from radar systems are available and care must be taken in that storing data at a trials site onto magnetic or paper tape, one is only postponing the actual job of data reduction. Accordingly a set of instrumentation equipment was designed and built especially for use with scanning radars, brief details of which are given in Table 3.

Each of the channels is driven from an accurately positioned digital range and azimuth marker so that the

Table 3
Details of instrumentation

No. of channels available	4
No. of discrete sampling levels	8
No. of pulses countable within a beam dwell time	2048
Sampling time in each channel	50 n s

effective amplitude samples are within the pulse length used for the number of pulses contained in the azimuthal extent of the azimuth gate. The gate dimensions were set to encompass several nominal beamwidths of the higher frequency channel and more than the 3 dB beamwidth of the 10 cm carrier.

Whilst the beamwidth concept and constraint has real meaning for point targets, in the case of extended or distributed clutter, particularly the more homogeneous clutter due to rain compared to sea or land, the count from the amplitude statiscizers would continue indefinitely. Care must therefore be taken in the interpretation of the results presented later in the paper.

7 Target Results

The aerial beams on the dual channel radar were aligned in azimuth to ensure target overlap from the signals received in the two channels, but the nose of each of the beams were displaced vertically to give the most efficient cover of the upper and lower air space on a single channel basis with dual cover from the horizontal to +10° where the beams overlap.

As a result the free space forecasts made in Section 4 apply for targets appearing on the nose of each beam and the general vertical coverage is determined by the cover of either or both carrier frequency depending on the method used for post receiver detectors signals combination. Figure 3 shows the relationship of the two beams on a range/height chart for a flat earth (as is sufficient for the short range shown).

These aerial arrangements must affect both target and clutter returns and the rest of the paper deals with results obtained from this particular equipment, and simple comparisons of sea clutter returns and small surface

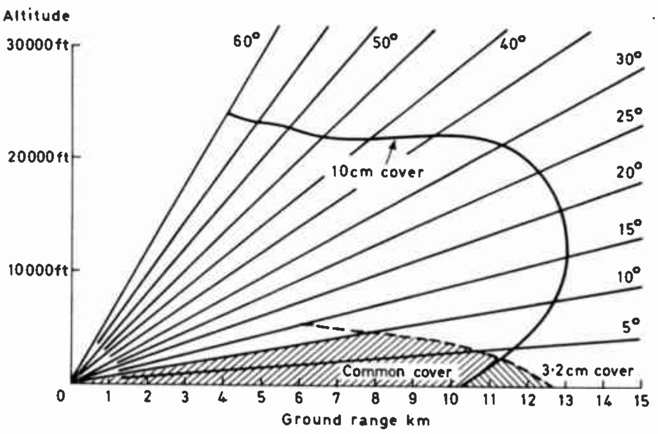


Fig. 3. Forecast coverage on target of r.c.s. $\sigma_{50} = 1$ m^2.

ships must be corrected for the gain reduction of the 10 cm beam on the water. Even so, the significant returns from shipping do allow a measure of diversity improvement to be obtained.

7.1 Photographs

Figures 4(a) and (b) illustrate the radar p.p.i. on the high and low carrier frequencies respectively, and Fig. 4(c) is a composite view of both channels on one display. As is customary for p.p.i. operation, the low dynamic video range merges both channel returns into a single optical return and it is not easy to distinguish a high amplitude, high frequency carrier return from a low frequency carrier, low amplitude, return as both tend to produce echoes of similar azimuth extent.

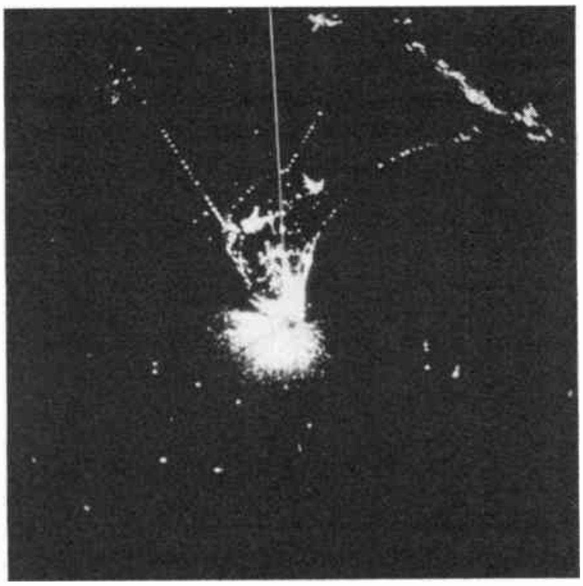
For this reason, the amplitude statistics in the next Section are more revealing as they are derived from each channel before video combination takes place.

7.2 Statistical Analysis of Each Pulse Return—Surface Targets

The instrumentation allowed the azimuth counting gate to be positioned on a chosen target whilst scanning and a count taken of the number of pulses reaching each of the staticizer levels ('hits'). The increment between steps was 5 dB for the results shown, so that the 'top score' of 8 was set some 35 dB above a threshold sufficiently above system noise in each receiver to have a substantially zero false alarm rate for the experimental time involved. On the curves the threshold crossing number is given on the Y axis as well as each data point.

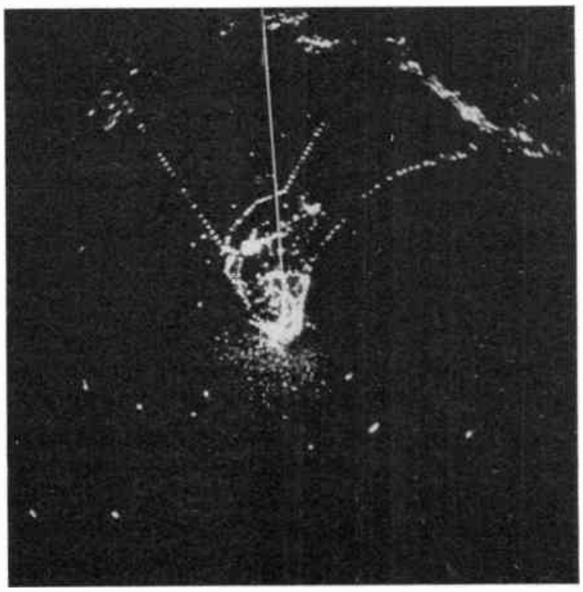
The first of three results is for a channel marker buoy some 3·75 nautical miles South of the radar carrying a large corner reflector of 150 m^2 peak at 3·2 cm (15 m^2 at 10 cm). This frequency sensitive factor, coupled to the fact that the 10 cm lobe is lifting well clear of the sea, justifies a median performance advantage of 17 dB to the high frequency carrier element of the system; one result is shown in Fig. 5(a). Because of the 'chance' nature of the peak returns the 50% value from each result gave an average of 18 dB. Numbers on the Y axis and the curves refer to the actual count value at each level. The next group of results are from a set of 10 scans taken from each of three ships at ranges of 5, 2 and 3·45 n. mi, these show the advantage varying from +7 dB to −6 dB for the low frequency carrier; furthermore, the shape of the curves change considerably and included an abrupt truncation of 3·2 cm readings in Fig. 5(d) with an extremely long-tailed result on what is already a log presentation of results in the same figure for the 10 cm carrier.

The final figure, Fig. 5(e), in this series is a pair of curves for a cumulative count for a longer range target at 8·5 n. mi taken over 10 aerial revolutions, where only a weak count occurred for the opportunity in each beam dwell time. The advantage to the 3·2 cm carrier channel is 18 dB averaged over 10 revolutions of the aerial (half a



(a)

(b)



(c)

Fig. 4.
(a) Surface cover on land and sea with 3·2 cm carrier. Range 12 n.mi. (b) Partial cover on land and sea with 10 cm beam set at +20° above horizontal. Note vestigial cover of high ground to N.E. of North marker. Range 12 n.mi. (c) Composite photograph after processing. Range 12 n.mi Single scan exposure.

Table 4
Summary of marine target trials (Plots shown for *)

Range n. mi	Target Description	Median Value of S>X dB per Trial
3·75	CS1 Buoy*	−17
3·75	,, ,,	−16
3·75	,, ,,	−20
2·00	Ship	+ 7
3·45	Ship	− 6
5·00	Ship	+ 3
8·5	Ship	−17

* CS1 Buoy is 'Channel Separation Buoy No. 1.

minute of sample observations).

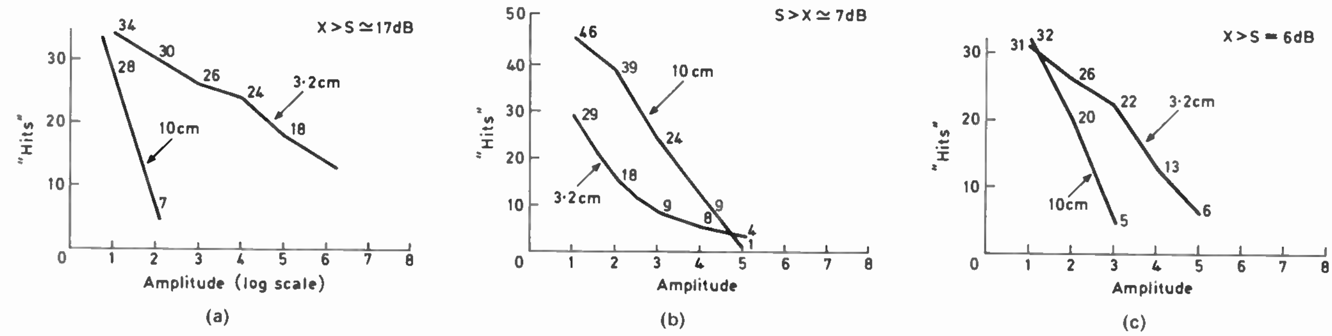
The problems of modelling surface radar performances against real surface targets is thus clearly shown and summarized in Table 4. These short term fades are as complex as the longer term fades separately reported by Williams[23] in observations from a higher aerial site over longer periods against a wide variety of shipping targets. Note how the shape of these probability density curves changes from concave to converse through straight, thus showing a wide range of short term fading models. We now turn to trials carried out with a target well clear of the Earth's surface and hence inside the transition range of both radars.

7.3 Target Results from Balloon Trials

To examine the relative gain of each carrier at various elevation angles, balloons covered with a 'Chemring' reflective coating were used as a control target.

As the coating mesh cover constrained the balloon for the whole of each run, the r.c.s. remained substantively constant for each trial at each altitude. An inflated diameter of 1·1 m theoretically gave both a peak and a median r.c.s. value of 1 m^2 for all these trials.

Figures 6(a) to (c) give the relative performance of the two carriers on the balloon target at 3 elevation angles. The interesting result is that, no doubt due to the differing beamwidths on the two carriers, the low-frequency, wide-beam carrier has in general more pulses or hits per beamwidth, but the higher-gain, high-frequency carrier aerial in all cases scores higher peak



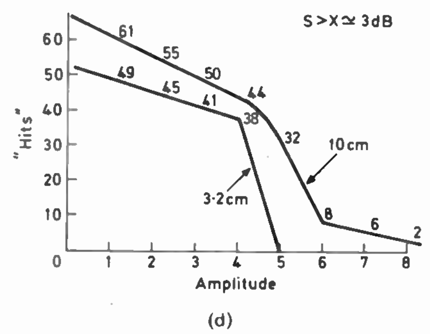
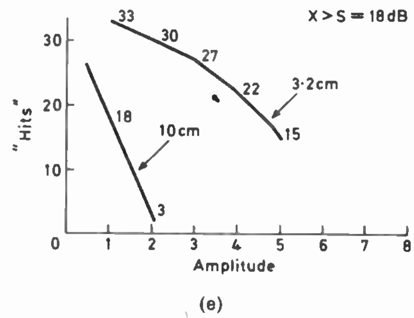


Fig. 5. (a) CSI buoy at $3\frac{3}{4}$ n.mi. (b) Ship at 2 n.mi. (c) Ship at 3·45 n.mi, cumulative for 10 aerial revolutions. (d) Ship at 5 n.mi, cumulative count for 10 aerial revolutions. (e) Ship at 8·5 n.mi, cumulative count for 10 aerial revolutions.



amplitude returns. Each of these results shown are the average of eight separate runs on the balloon for each elevation and all trials were carried out at a range of half a mile.

From further statistics taken on the tethered balloon it was noted that the median relative performance of the two carriers comes very close to theory for the individual results at 2°55′ elevation, and in all cases by extrapolating to a count of zero the 3·2 cm peak return exceeds all the 10 cm results.

All of the balloon trials discussed so far took place over a shingle ground plane which one would have expected to have a much lower microwave reflection coefficient than a smooth calm sea, and whilst it can be argued multipath nulls on either carrier could contaminate results, the null spacing, if present, would be of the order of less than 1° on the 10 cm carrier and three times as fine for the 3·2 cm carrier. In the wind, the balloon appeared to be surging up and down about 1° so that the null pattern would tend to be smoothed out over the number of independent readings taken.

In the time available only one ballon was released for free flight and this was followed out in range further on the 10 cm carrier than the high frequency carrier. Unfortunately, the mist and cloud over the sea came in and precluded any theodolite elevation measurements at that range, but as the balloon faded from sight at about 30° from the horizontal, it must be assumed that it had risen above the 3·2 cm beam quite early in the flight.

8 Rain Clutter

An estimate of the radar cross-section of medium rain at 10 km has already been made in the description of the equipment. For a rainfall rate of 4 mm per hour, the high-frequency carrier, even with its superior spatial resolution, has an r.c.s. of nearly 10 m^2 compared to less than half a square metre on the lower frequency.

Figure 7 shows a photograph of the p.p.i. with extensive echoes from frontal rain near and over the site on 31st January 1980, as observed on the high-frequency carrier.

The expected rain power returned on the 10 cm channel is reduced by the ratio of $[\lambda_s/\lambda_x]^4$ offset by the greater angular beam dimension of the 10 cm channel. These factors yield an overall reduction of 12·8 dB and are set out in column 5 of Table 5.

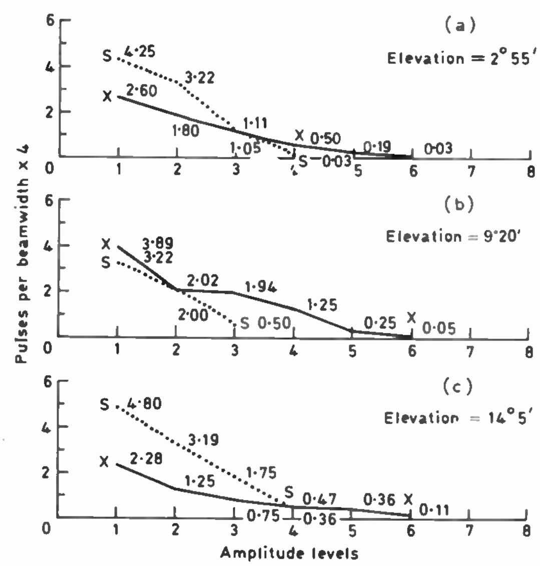


Fig. 6. Relative performance with elevation.

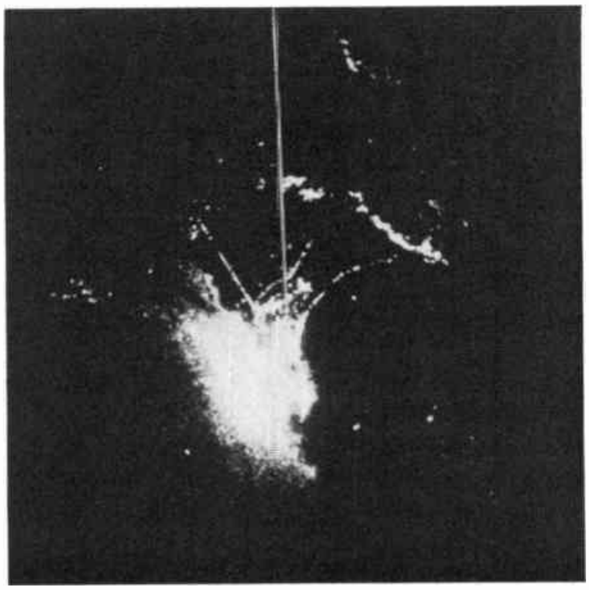


Fig. 7. P.p.i. photograph with rain over the site (3·2 cm carrier).

Table 5

Summary of rainfall clutter measurements compared to theoretical estimates

Carrier	Pulse Length µs	Count at Peak Level above Level 1	Measured Median Level. above Level 1	Calculated R.C.S. of Rain Clutter Expected with respect to 1 µs at λ = 10 cm
3·2 cm	1 µs	63	10 dB	0 dB
3·2 cm	0·25 µs	65	+8dB	−4 dB
3·2 cm	0·05 µs	60	+6 dB	−13 dB
10 cm	1 µs	0	—	−12·8 dB
10 cm	0·25 µs	0	—	−16·8 dB
10 cm	0·05 µs	0	—	−25·8 dB

The staticizer, however, counts threshold crossings in each of the possible eight levels and thus the count is independent of pulse length being purely peak conscious. One of the three sets of counts summarized in Table 5 is shown as Fig. 8 and the 50% count indicates that the median value of the distribution lies at level 3 which is 7 dB up on the lowest level. All the distributions tend towards normal shape on logarithmic scales.

9 Sea Clutter

The unwanted returns from all but a flat calm sea have provided a source of target screening and false alarm since the first use of radar. The literature is prolific and to some extent contradictory. This paper reports a series of measured results taken with the dual channel radar on a single day.

Figure 4 has already shown two examples of sea clutter return and Fig. 9 shows the combined effect of rain and sea clutter received on both carriers as a rainstorm passed over the site. Obstructions to the East partly obscure the radar aerials so no great significance should be attached to the partially blind sector on the right-hand side. The radial heading marker is approximately 20 n.mi long, from which the extent of clutter may be judged.

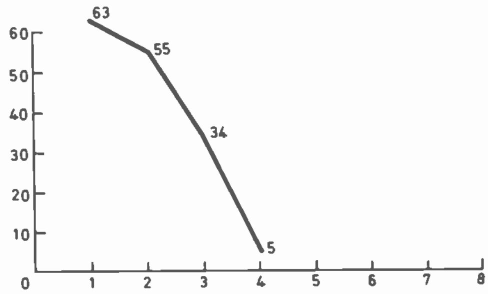


Fig. 8. Distribution of threshold crossings for rain returns taken over 10 aerial revolutions on X-band.

Partly because of the much lower backscatter coefficient of the sea surface for a wavelength of 10 cm and partly because of the upward tilt of the 10 cm aerial beam and also because of the shorter transition range and low aerial height above the sea, backscatter from the surface is significantly reduced on 10 cm compared to the 3·2 cm carrier, even though the horizontal beamwidth is twice as wide.

9.1 Sea Clutter—Photographic Evidence

During the day of 31st January 1980 high winds created sufficient waves in the Channel as to exhibit white tops—sea state 3 to 4—with the wind gusting up to force 6 or 8 from the S.W. Figures 10(a) and (b) illustrate the extent of sea clutter returns just prior to the amplitude statistics being taken, the radial marker on these photographs points due North so that the amplitude runs taken on a bearing of 226° were near upwind to give large sea clutter returns. (The light clutter returns are in fact due to rain South of the site.)

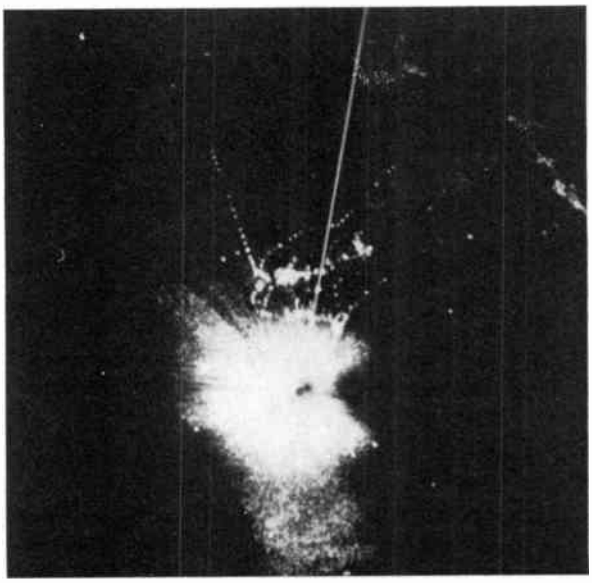
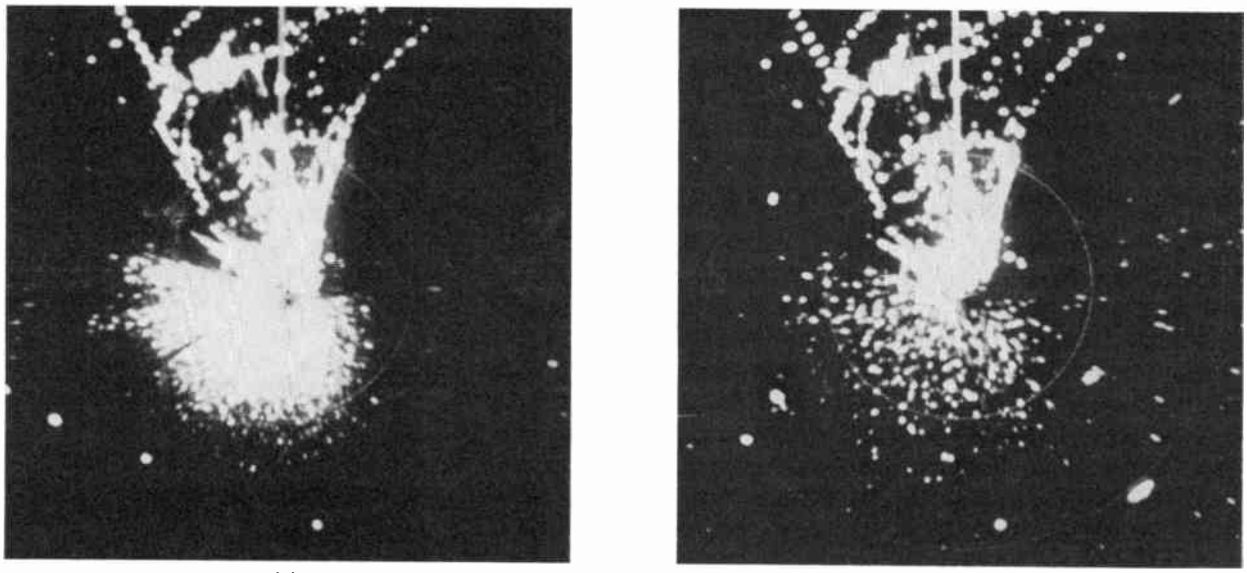


Fig. 9. General view of rain and sea clutter obscuring the area around the radar site on 24 n.mi range during a rainstorm. 31st January 1980.



(a) (b)

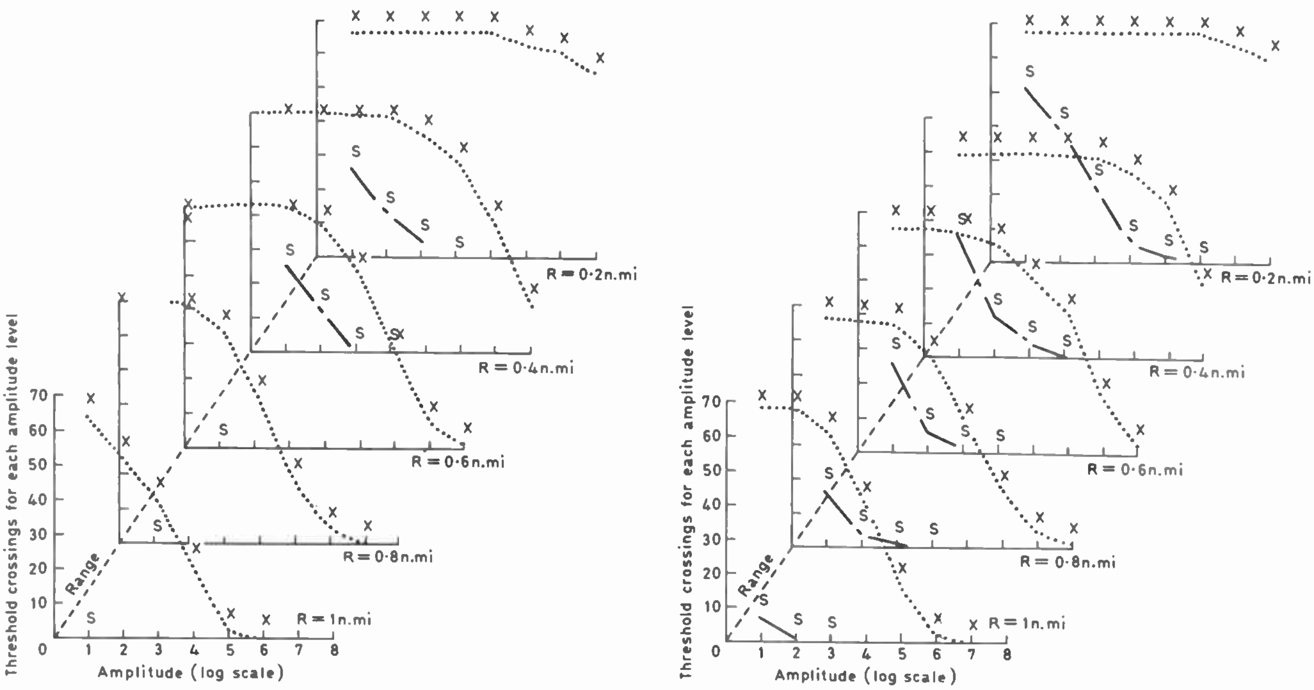
Fig. 10. (a) P.p.i. with 8 n.mi radius showing 4 n.mi of sea clutter and half a dozen ships. (b) 8 n.mi p.p.i. with sea clutter residue and shipping targets, taken later in the day.

9.2 Sea Clutter—Statistics

During the period of high seastates, (31.1.1980), various observations were made as summarized in Table 6. All the results were taken in a period of 4 hours, during this time the sea appeared to be consistently rough but in the absence of proper instrumentation this is only an observer's opinion.

The first series of curves (Fig. 11) are two sets of range runs starting at 0·2 n.mi (clear of the sea shore) and these are shown as Figs. 11(a) and (b). These two runs in fact were separated by two hours to sample the sea returns over a long period with a 1 μs pulse.

The next series (Fig. 12) is at the same ranges but with a pulse of 0·25 μs and to complete this range dependent count, Fig. 13 is for a pulse length of 0·05 μs when it will be seen there is a significant reduction of threshold crossing counts.



(a) (b)

Fig. 11. (a) Sea clutter amplitude statistics with range as a parameter. Pulse length = 1 μs. (b) Sea clutter amplitude statistics with range as a parameter. Pulse length = 1 μs.

Table 6

Summary of sea clutter statistics taken on 31st January 1980 on both 10 cm and 3·2 cm carriers

Run Ref. Nos.	Common Pulse Length	Range n.mi	Feature of Each Set of Runs
9 to 37	1 μs	0·2 to 3	Threshold crossings in each level.
38 to 57	1 μs	0·2 to 2·1	Threshold crossings in each level.
60 to 71	1 μs	0·2 to 2·4	Threshold crossings in each level.
72 to 83	0·25 μs	0·2 to 2·4	Threshold crossings in each level.
84 to 88	0·05 μs	0·2 to 1·2	Threshold crossings in each level.
90a to j	0·05 μs	1 only	Time series at a single range
91a to j	0·25 μs	1 only	Time series at a single range
92a to j	1·0 μs	1 only	Time series at a single range
93/1 to 10	1·0 μs	0·4 only	Time series at a single range
94/1 to 10	0·25 μs	0·4 only	Time series at a single range
95/1 to 10	0·05 μs	0·4	Time series at a single range
96/1 to 7	0·25 μs	0·4	Cumulative count of crossings on each level at single range

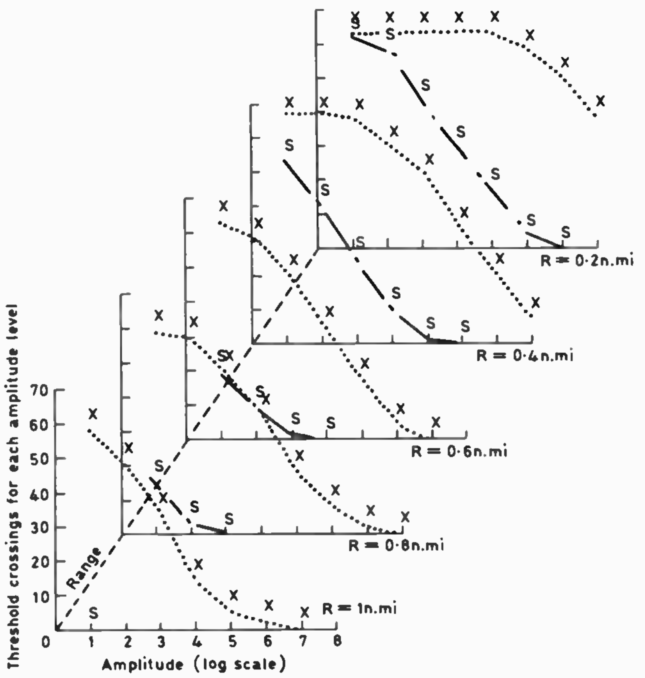


Fig. 12. Sea clutter amplitude statistics with range as parameter Pulse length = 0·25 μs.

9.3 Discussion and Further Analysis of Data

It is usual to consider the sea clutter return amplitude decreasing with reduction of pulse length but in these measurements the amplitude staticizer gate length is held constant and the transmitted pulse tends to 'flood light' across the gate in the range dimension, and we are probably only starting to enjoy inter-clutter visibility on the shortest pulse with the receiver I.f. bandwidth at 15 MHz.

Having examined the range and pulse length dependency of the sea clutter amplitude threshold crossing, we will now examine a more comprehensive series of plots (Fig. 14) where the y axis is once again a threshold crossing, but this time the x axis is ranged from

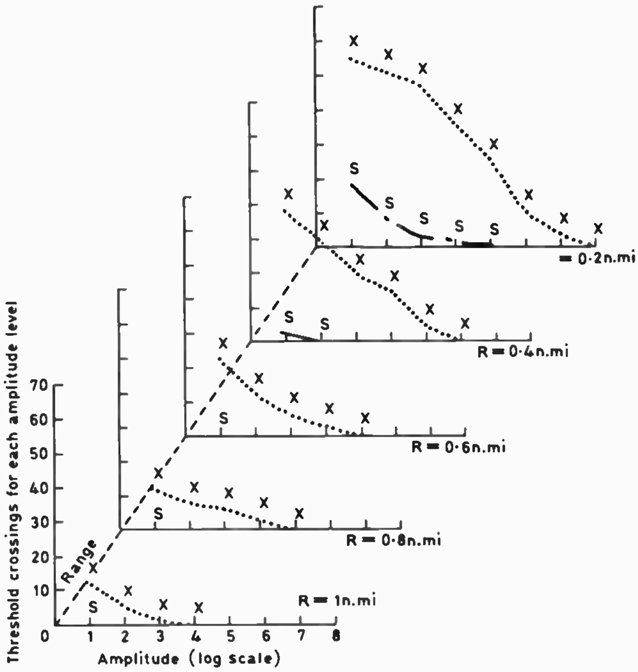


Fig. 13. Sea clutter amplitude statistics with range as a parameter. Pulse length = 0·05 μs.

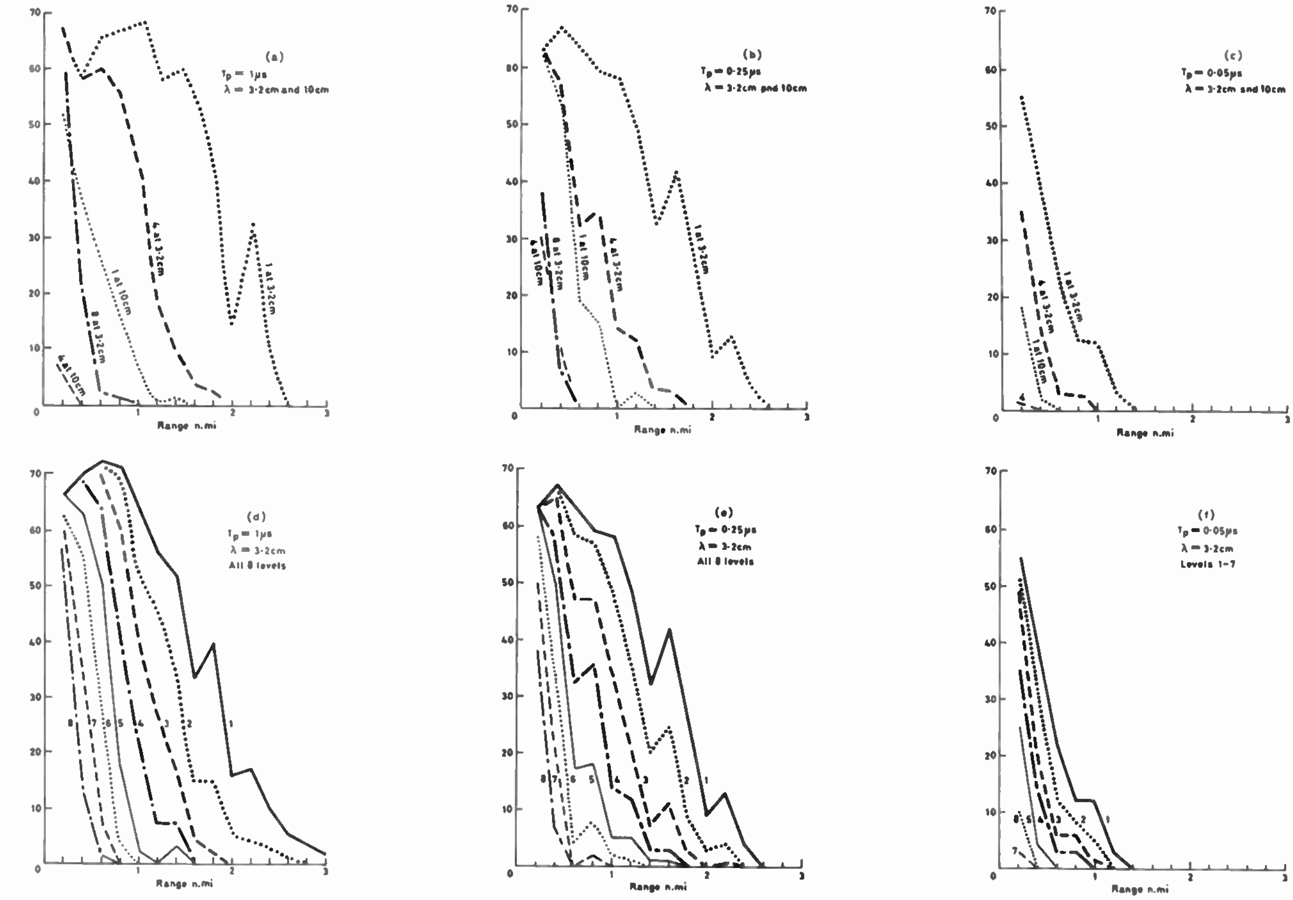
Table 7.

Summary of Figs 14(a) to 14(g)

λ	Fig.	Source Data	Pulse Length	Level Shown
Both	(a)	60 to 71	1 μs	1–4–8
Both	(b)	72 to 83	0·25 μs	1–4–8
Both	(c)	84 to 89	0·05 μs	1–4
3·2	(d)	9 to 37	1 μs	1 to 8
3·2	(e)	72 to 83	0·25 μs	1 to 8
3·2	(f)	84 to 89	0·05 μs	1 to 7
10	(g)	72 to 83	0·25 μs	1 to 6

0·2 to 3 n.mi. Each figure contains several plots, one for each amplitude, summarized in Table 7.

The interesting features of sea clutter amplitude statistics are quickly submerged by any sort of averaging process. To illustrate this two sets of cumulative threshold crossing counts in each of the 8 amplitude levels of the staticizer are given in Fig. 15(a) (for the X-band channel) and Fig. 15(b) for the S-band channel, both recorded at a range of 0·4 n.mi. Over the 4 min period each sample from 10 aerial revolutions shows an orderly increase on the previous set, so that the (vertical) distribution of amplitude count is substantially the same as time progresses. Note, however, the closeness of the count for levels 1 and 2, a feature which has appeared although displayed in a different form in the shorter range data of Figs. 11–12 for the X-band channel using 0·25 μs and 1 μs pulse lengths where the crossing count



appears to be saturated at the lower amplitude levels.[27].

In these cumulative results the general increase of sea clutter return for X-band over S-band is again clearly demonstrated, in fact the higher amplitude levels 6 to 8 are never reached at all on the S-band channel.

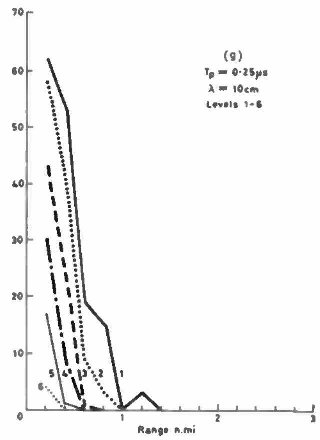


Fig. 14. Sea clutter statistics.

If now the statistics of an individual set of data points on Fig. 15(a) or (b) are expanded to resolve their fine structure in time, the presentation is far less orderly. Within a minute of the data shown in Fig. 15 being recorded, the final sea clutter measurements were made at the same range, bearing and pulse length. Figure 16(a) on X-band and Fig. 16(b) on S-band showed such a confused situation that level 2 was omitted for clarity and levels 3 and 4 only partly shown. For the case of S-band the failure to reach levels 5 to 8 in any of the 10 scans recorded produced a slightly more orderly diagram but the danger of designing constant false alarm rate (c.f.a.r.) circuits from the highly smoothed data shown in the preceding two diagrams is clearly shown.

In comparing these last two figures some temporal correlation of peaks and troughs is observed, but the trough at scan 4 in Fig. 16(a) seems to be anticipated in Fig. 16(b). The conclusion is that the peak to median ratio of sea clutter amplitudes experienced on each scan do vary a great deal and good fits to statistical models such as Rayleigh, log normal, Weibull or log Weibull, is only obtained for data smoothed over many scans, thus averaging out sea surface short term phenomena which is shown on any real time radar display.

10 Composite Picture

The provision of two radar channels presents a problem in how best to initially set up each channel, then how to mix them for use on one display. Figure 17(a) is the p.p.i. result of taking each channel, first applying c.f.a.r. processing, then combining the results. The 12 n.mi radius picture gives plenty of land and coastal features together with plenty of shipping targets. Figure 17(b) is the result of c.f.a.r. filtering being applied to the 3·2 cm carrier channel only, which is then combined with the

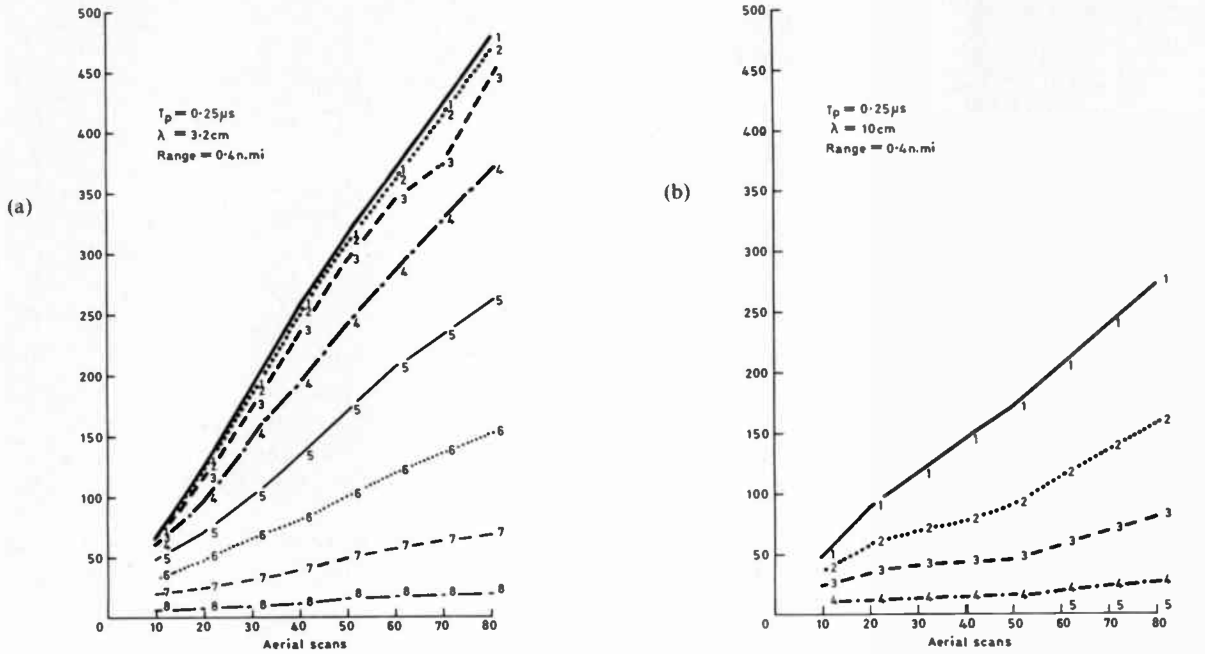


Fig. 15. (a) Cumulative count of sea clutter statistics for all 8 levels (b) Cumulative count of sea clutter statistics for 5 levels.

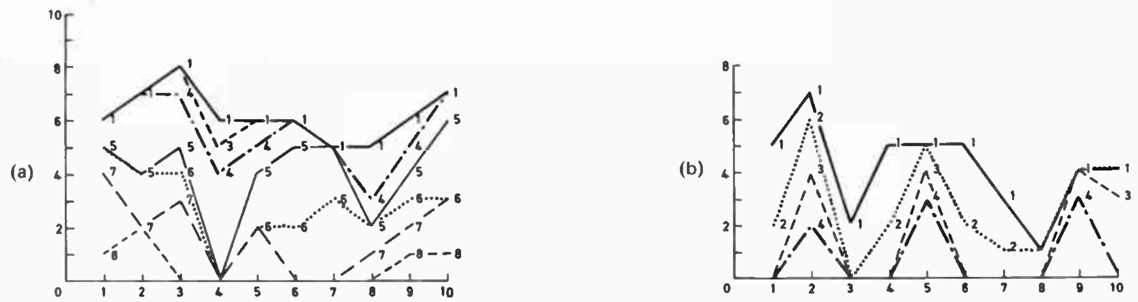


Fig. 16. Sea clutter statistics for 3.2 cm and 10 cm as a timehistory of 10 separate scans all at 0·4 n.mi. (a) $\lambda = 3{\cdot}2$ cm. $T_p = 0{\cdot}25\,\mu$s. (b) $\lambda = 10$ cm. $T_p = 0{\cdot}25\,\mu$s.

10 cm carrier. This shows a little more light sea return from the 10 cm channel and represents the best detection compromise.

From the final series of pictures taken during a heavy rainstorm a few days later and showing the obscuration caused by rain clutter on the 3·2 cm channel only Fig. 18 is shown, giving the results of incorrect use of c.f.a.r. following video combinations.

11 Conclusions

This paper has shown how the correct use of a twin-carrier system can provide superior performance to that of a single-channel radar and will be far more cost-effective in meeting a wide range of operational requirements than a single-carrier set, even with 3 dB more free space performance due to both diversity gain[2,3] and optimum c.f.a.r. processing of sea and rain clutter.

This is brought about by the essentially different fading pattern of targets in the critical surface and lower air space where frequency diversity is very effective. In the higher elevations, the freedom from both sea and rain clutter on 10 cm, coupled to the broad vertical beam of the 10 cm wavelength, '12 ft' aperture aerial, provides good air cover.

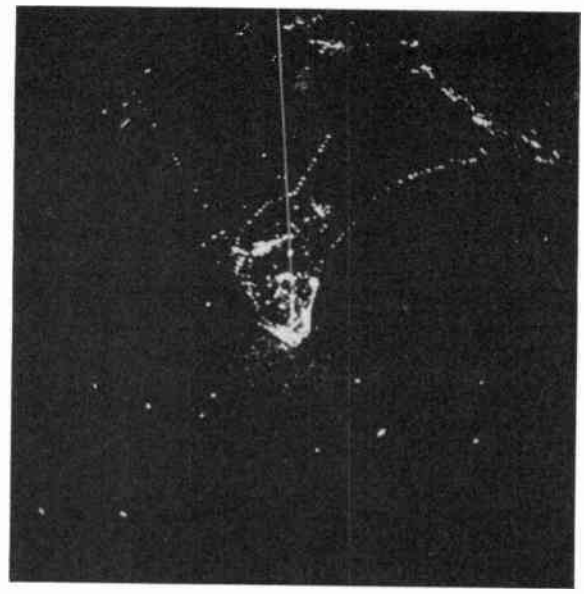
12 Acknowledgments

The author would like to express his sincere thanks to Racal–Decca Marine Radar Ltd for the support given in all the trials work.

13 References

1 Skolnik, M. I., 'Introduction to Radar Systems' (McGraw-Hill, New York, 1962)
2 Barton, D. K., 'Radar System Analysis' (Prentice Hall, Englewood Cliffs, N.J., 1964)
3 Nathanson, F. E., 'Radar Design Principles' (McGraw-Hill, New York, 1969)
4 Skolnik, M. I. (Ed.), 'Radar Handbook' (McGraw-Hill, New York, 1970)
5 Long, M. W., 'Radar Reflectivity of Land and Sea' (D. C. Heath, Lexington, MA, 1975)
6 Barton, D. K. (Ed.), 'Radar Clutter'. (Artech, Dedham, MA, 1975)
7 'Radar—Present and Future', London 1973. (IEE Conference Publication No. 105)

(a)



(b)

Fig. 17. (a) 12 n.mi p.p.i. with both carriers combined after individual c.f.a.r. filtering. (b) 12 n.mi p.p.i. similar to above but with c.f.a.r. in 3·2 cm carrier only.

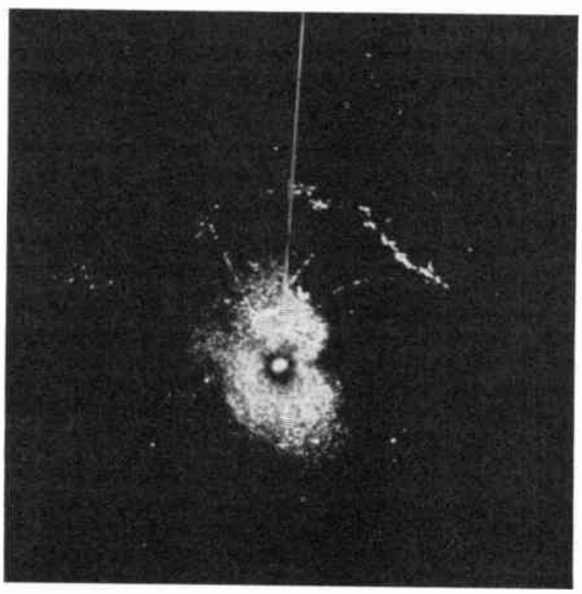


Fig. 18. Incorrect processing. 24 n.mi in rain.

8 International Radar Conference, 1975. (IEEE 75CH0938—1AES.)

9 'Radar—77', London 1977. (IEE Conference Publication No. 155.)

10 'The Sea Surface, Target or Clutter?' (IEE, Colloquium London, 12th February 1979. (Abstracts available)

11 Valenzuela, G. R., 'Scattering of electromagnetic waves from the Ocean', in 'Surveillance of Environmental Pollution and Resources by EM Waves'. (Ed.) Terje Lund. (D. Reidel, Dordrecht 1978.)

12 Williams, P. D. L., Cramp, H. D. and Curtis, K. 'Experimental study of the radar cross-section of maritime targets,' *IEE J. Electronic Circuits and Systems*, **2**, no. 4, pp 121–35, July 1978.

13 Valenzuela, G. R., 'Theories for the interaction of electromagnetic and ocean waves—a review,' *Boundary-Layer Meteorology*, **13**, pp. 61–85, 1978.

14 Daley, J. C., *et al.*, 'Sea Clutter Measurements on Four Frequencies', NRL Report No. 6806, Washington, 1968.

15 Sittrop, H., 'X-band and Ku-band radar backscatter characteristics of sea clutter', Proc. URSI Specialist Meeting on Microwave Scattering and Emission from the Earth, Berne, September 1974.

16 Dyer, F. B., Currie, N. C. and Applegate, M. S., 'Radar backscatter from land, sea, rain and snow at millimetre wavelengths', Radar 77, London 1977, (IEE Conference Publication No. 157)

17 Lewis, B. L. and Olin, I. D. 'Experimental study and theoretical model of high-resolution radar backscatter from the sea', *Radio Science* **15**, no. 4, pp. 815–28, 1979.

18 Williams, P. D. L., 'Limitations of radar techniques for the detection of small surface targets in clutter', *The Radio and Electronic Engineer*, **45**, no. 8, pp. 379–89, August 1975.

19 Vannicola, V., 'Detection of slow fluctuating targets with frequency diversity channels', *IEEE Trans. on Aerospace and Electronic Systems* **AES–10**, no. 1, pp. 43–52, January 1974.

20 Poelman, A. J., 'Comments on "Detection of slow fluctuating targets with frequency diversity channel"', *IEEE Trans.* **AES–10**, no. 4, pp. 544–5, July 1974.

21 Barton, D. K. (Ed.), 'Frequency Agility and Diversity'. Vol. 6 of "Radars" by (Artech, Dedham MA)

22 Barton, D. K., 'Comments on the citing of restricted reports', *IEEE Trans.*, **AES–10**, no. 5, pp. 727–8, September 1974.

23 Williams, P. D. L., 'Medium term fading of radar targets at sea with special reference to operation at 3 cm and 10 cm,' *Proc. Instn Elect. Engrs*, **127**, pt. F, no. 3, pp. 212–24, June 1980.

24 Williams, P. D. L., 'Change of the fading characteristics of reputed steady targets as a function of sea state in a marine environment', *Electronics Letters*, **6**, no. 26, pp. 853–5, 31st December 1970.

25 Sekine, M., Mushe, T., Irabu, T. *et al.*, 'Non-Rayleigh weather clutter', *Proc. Inst Elect. Engs*, **127**, pt. F, no. 6, pp. 471–4, December 1980.

26 Maaløe, J. and Williams, P. D. L., 'Weibull distribution of sea clutter: correction of marine radar data to improve the fit. (To be published.)

27 Williams, P. D. L. and Maaløe, J. 'Radar sea clutter: remaining deviations from a Weibull model. (To be published.)

Manuscript first received by the Institution on 17th July 1980 and in revised form on 3rd March 1981
(Paper No. 2113/AMMS 107)

UDC 537.211 : 531.112.8
Indexing Terms: Electric fields, Relativity

Electric forces

Professor R. A. WALDRON, M.A., Sc.D., F.Inst.P., F.I.M.A., C.Eng., F.I.E.R.E.*

SUMMARY

In place of Coulomb's Law,

$$F = q_1 q_2 / 4\pi\varepsilon_0 r^2,$$

the formula

$$F = (q_1 q_2 / 4\pi\varepsilon_0 r^2)\sqrt{1+v^2/c^2}$$

is proposed, where v is the velocity of one charge relative to the other. Using this velocity-dependent force law, it is shown how Ampère's law can be derived; Maxwell's equations can then be obtained, and their limitations can be discussed. The concept of field is shown to be valid only in the limiting case of vanishing velocities, so that Maxwell's theory is approximately valid only when any motion is slow and any rate of change of field is small. When these conditions do not hold, Maxwell's theory is invalid, but the velocity-dependent force law can be used. The motions of charged particles in electric and magnetic fields are discussed. Anomalous apparent decay times of fast mesons are attributed not to time dilation but to speeds greatly exceeding c. Some anomaly remains, and it is suggested that new experimental observations are necessary to resolve this.

* *School of Mathematics, Ulster Polytechnic, Shore Road, Newtownabbey, Co. Antrim BT37 0QB.*

Principal Symbols

E, H, D, B	electric and magnetic fields and inductions
J	current density
q	electric charge
q_v	volume density of charge
q_e	electronic charge
ε_0, μ_0	permittivity and permeability of vacuum
F	force
x, t	space and time co-ordinates
w	relative velocity of two co-ordinate systems
v	velocity of charged particle according to the theory proposed in this paper
u	velocity of charged particle according to Maxwell–Lorentz theory
m	mass
m_e	electronic mass
r	distance between two point charges; radius of curved path of charged particle in magnetic field
ρ	radius of curved path of charged particle in radial electric field
V	voltage
W	work
i	current
δl	element of length of conductor
c	velocity of light in vacuum, relative to source at time of emission
τ	decay time of μ mesons

1 Introduction

For over a hundred years Maxwell's equations have been the basis of a large part of electrical science and electromagnetism. Shortly after the publication of Maxwell's electromagnetic theory discharges in gases were being studied, and it was in 1876 that Goldstein gave the name 'cathode rays' to certain beams, the nature of which was unknown. Several measurements of the ratio of charge to mass were made towards the end of the nineteenth century, and in 1899 J. J. Thomson measured both this ratio and the charge. Measurements of the ratio were made by observing the curvature of the track in a magnetic field. Previously, in 1897, Thompson had also observed the deflection in an electric field, and so had been able to measure the charge-to-mass ratio and the velocity, which turned out to be in the neighbourhood of one-tenth of the velocity of light, c.

The discovery of the electron is attributed to Thomson in 1897, but it had been predicted some time earlier and the name 'electron' was coined by Johnstone Stoney in 1891. H. A. Lorentz had developed his theory of electrons, which was based on Maxwell's equations, and according to which the mass of the electron was of electromagnetic origin.

Maxwell's equations accounted for forces between electrically charged and/or current-carrying bodies in terms of electric and magnetic fields, but did not cover the motions of highly charged rapidly moving particles

0033–7722/81/110553+08 $1.50/0

such as electrons in electric and magnetic fields. The forces on such particles were expressed in another formula arising from the theory of electrons, known as the Lorentz force formula. Thus at the end of the nineteenth century the whole of electrical and magnetic science was based on Maxwell's equations,

$$\left.\begin{array}{ll}\text{(a)} & \operatorname{Div} \mathbf{D} = q_v \\ \text{(b)} & \operatorname{Div} \mathbf{B} = 0 \\ \text{(c)} & \operatorname{Curl} \mathbf{E} = -\partial \mathbf{B}/\partial t \\ \text{(d)} & \operatorname{Curl} \mathbf{H} = \partial \mathbf{D}/\partial t + \mathbf{J}\end{array}\right\} \quad (1)$$

and the Lorentz force formula

$$\mathbf{F} = q(\mathbf{E} + \mathbf{u} \times \mathbf{B}) \quad (2)$$

for the force $\mathbf{F}$ on a charge q moving with velocity $\mathbf{u}$ in an electric field E and a magnetic field $\mathbf{B}$.

Maxwell's equations expressed relations between electromagnetic fields whose meaning lay in stresses and strains in the aether. Lorentz's theory of electrons preserved these notions, and the velocity $\mathbf{u}$ of equation (2) is velocity with respect to the aether. Other consequences of the theory were the well-known Lorentz–FitzGerald contraction and time dilation and the Lorentz transformations.

Still to-day electromagnetism and electrodynamics continue to be based on equations (1) and (2), although the failure to detect the aether experimentally led to a disenchantment with the metaphysical basis of the theory at the turn of the century. The minds of physicists were then receptive to Einstein's so-called special theory of relativity which is, as is explained by the writer elsewhere (ref. 1, pp. 40–42), nothing else than Lorentz's theory of electrons under a different guise. Its purpose, in which it was successful, was to save the Maxwell–Lorentz theory whose essence is equations (1) and (2). The cost was the preservation of such notions as the Lorentz–FitzGerald contraction and time dilation, and numerous contradictions which have been pointed out by various writers [ref. 1, chap. 3; refs. 2–10]; the Lorentz transformations,

$$\left.\begin{array}{ll} x' = \dfrac{x - wt}{\sqrt{1 - w^2/c^2}}; & t' = \dfrac{t - wx/c^2}{\sqrt{1 - w^2/c^2}}; \\ x = \dfrac{x' + wt'}{\sqrt{1 - w^2/c^2}}; & t = \dfrac{t' + wx'/c^2}{\sqrt{1 - w^2/c^2}}; \end{array}\right\} \quad (3)$$

are also at the core of the theory. Here x' and t' are one space co-ordinate and the time co-ordinate of an inertial system S′ which is moving in the direction of the positive x axis of another inertial system S, in which the time co-ordinate is t, the relative velocity being w.

It is not intended to discuss relativity theory, as such, here, although it will be impossible to avoid some mention of it. The purpose of this paper is to propose a new theory of electromagnetism and electrodynamics. The significance of Maxwell's equations and the Lorentz force formula have not previously been appreciated, and it has not been understood that the range of their applicability has its limits. The new theory will be shown to reduce to the Maxwell–Lorentz theory where this is valid, and to provide a means of dealing with electrodynamic phenomena outside the scope of the Maxwell–Lorentz theory. Finally, the principles by which experimental tests of the theory might be sought will be discussed.

2. Limitations of the Maxwell–Lorentz Theory

When first formulated, Maxwell's equations were thought of as expressing relations between fields, which were strains in the aether. The fields were carried, as it were, by the apparatus (electrodes, coils, etc.) used to generate them. Thus if a system of electrodes and currents moved through the aether, the strains also moved, and this motion did not affect the relations between fields that were expressed by Maxwell's equations. This was tested in a famous experiment by Trouton and Noble[11] in which a capacitor, filled with a solid dielectric, was suspended and rotated into different attitudes, so that the motion of the Earth (and therefore of the capacitor) was in different directions with respect to the orientations of the capacitor plates. No anisotropy of the dielectric constant of the filling material was detected.

The conclusion drawn from this was that Maxwell's equations hold good, regardless of the motion of a system of field generators with respect to the aether. Fair enough, if the aether exists. If it does not, the inference to be drawn is that Maxwell's equations do not permit an absolute velocity to be detected. What must be appreciated here is that in any experiment so far performed to test the validity of Maxwell's equations, all parts of the system of field generators are at rest with respect to one another, or in motion with relative velocities minute in comparison with c, whatever may be the state of motion of the system as a whole.

Thus Maxwell's equations are valid only for systems having no relatively moving parts.

The Lorentz force formula is in accordance with observation for slowly-moving charged bodies. When the velocity u becomes so large that u^2/c^2 is appreciable compared with unity, however, the observations do not agree with equation (2) unless the mass m' of a charged body is taken to be

$$m' = m/\sqrt{1 - u^2/c^2} \quad (4)$$

where m is the mass at zero velocity. Equation (4) is predicted by Lorentz's theory of electrons, when u is the velocity of a charged body with respect to the aether, and by Einstein's relativity theory, when it is the velocity with respect to the observer. The observer may be interpreted as the field generators, and it must be understood that these are at rest with respect to one another. What is meant by a field when its generators are in relative motion is not at all clear, so the Lorentz force formula does not hold in this case, any more than Maxwell's equations. The force may, however, be treated by the methods of the new theory, if the necessary data are available to set up the problem.

In spite of the great amount of experimental evidence apparently agreeing with it, the Lorentz force formula can only be regarded as established for small velocities. The formula is itself used to determine the quantity u in many experiments, and whether or not u is truly the velocity depends on the validity of the formula. There are rare experiments in which the true velocity v of a particle is measured by timing it over a certain distance, but then the velocity is not measured electromagnetically. Thus v and u have never been compared, and while equation (2) is well established in terms of u, we do not in fact know that u is the velocity, except when it is so small that u^2/c^2 is negligible in comparison with unity.

Thus the Lorentz force formula can only be regarded as established for the case of a particle moving at vanishingly small velocity with respect to a system of field generators at rest with respect to one another.

The conflict between classical electromagnetism and Newtonian mechanics arises from confusion over the limitations of the former. It has been assumed valid for field generators with moving parts and for rapidly-moving particles. This has brought it into conflict with Newtonian mechanics. To resolve the conflict Einstein's theory of relativity has been formulated. Newtonian mechanics is replaced by Einstein's mechanics, with the Lorentz transformations, time dilation, velocity-dependent mass, and other nightmarish concepts, not to mention unresolvable contradictions which invalidate it as a contender for consideration as a physical theory. What is necessary instead is a reformulation of electromagnetism which reduces to the Maxwell–Lorentz theory for vanishingly small velocities, and agrees with Newtonian mechanics in general. Such a theory will be proposed in the next Section.

3 A Velocity-dependent Coulomb's Law

Coulomb's law,

$$F_E = q_1 q_2/4\pi\varepsilon_0 r^2, \tag{5}$$

for the electric force F_E between two charges q_1, q_2, separated by a distance r, was established for the case when the charges are at rest with respect to each other. It is conceivable that if one charge is moving with respect to the other, the force between them might depend on the relative motion. Assuming that the force depends only on their relative velocity v and not on their relative or absolute acceleration of higher-order derivatives of v, it has been found that the force law

$$F_E = (q_1 q_2/4\pi\varepsilon_0 r^2)\sqrt{1+v^2/c^2} \tag{6}$$

provides a basis for a development of electromagnetism and electrodynamics which does not conflict with Newtonian mechanics and is self-consistent. Here v is the magnitude of the relative velocity, independent of its direction relative to that of the line joining q_1 to q_2. ε_0 is not to be thought of as a property of vacuum; it is a constant introduced to make the units and dimensions balance correctly. The adoption of equation (6) does not mean that accelerations or higher derivatives of v may not be found to be necessary in the future, but this is a matter to be considered in the light of experimental evidence not yet available.

In this Section the essential principles of electromagnetism and electrodynamics will be developed from equation (6) and the results will be compared with those of the Maxwell–Lorentz theory.

3.1 The Electric Field

If the charge q_2 is at rest with respect to q_1, it experiences a force given by equation (5), Coulomb's law; this is obtained from equation (6) on putting $v = 0$. The quantity $q_1/4\pi\varepsilon_0 r^2$ is an electric field. If charge is distributed over surfaces or throughout volumes, the net effect at a point can be obtained by means of the superposition principle, integrating over the charge distributions. Thus electric field in the new theory means the same as formerly—it is a quantity such that when multiplied by q it gives the force on the charge q, provided that q is at rest with respect to the charges producing the field, i.e. with respect to the electrodes, as long as these electrodes are at rest with respect to each other. The field resulting from a system of electrodes can be designated $\mathbf{E}$, and if $\mathbf{v}$ is the velocity of a point charge q relative to a co-ordinate system S in which all the electrodes generating $\mathbf{E}$ are at rest, the force on q is, analogously to equation (6),

$$\mathbf{F}_E = q\mathbf{E}\sqrt{1+v^2/c^2} \tag{7}$$

regardless of the direction of $\mathbf{v}$ relative to that of $\mathbf{E}$.

3.2 Motion of a Charged Particle Perpendicular to an Electric Field

We consider a pair of concentric cylindrical electrodes, producing a radial field varying radially as $1/\rho^2$, ρ being the distance from the common centre of the electrodes. An electron is injected so as to move in a transverse plane; if its azimuthal velocity v is correct, it will move in a circular arc with constant velocity under a constant field $\mathbf{E}$. The force on the electron will be

$$\mathbf{F}_E = q_e\mathbf{E}\sqrt{1+v^2/c^2}$$

and it will have an acceleration v^2/ρ towards the centre. Thus

$$q_e E\sqrt{1+v^2/c^2} = m_e v^2/\rho \tag{8}$$

or

$$E\rho = \frac{m_e v^2/q_e}{\sqrt{1+v^2/c^2}}. \quad (9)$$

According to the Maxwell–Lorentz theory, the force is $q_e E$, the acceleration is u^2/ρ, and the mass is $m_e/\sqrt{1-u^2/c^2}$, so that

$$q_e E = \frac{m_e u^2/\rho}{\sqrt{1-u^2/c^2}} \quad (10)$$

or

$$E\rho = \frac{m_e u^2/q_e}{\sqrt{1-u^2/c^2}}. \quad (11)$$

Equations (9) and (11) agree if

$$v^2/\sqrt{1+v^2/c^2} = u^2/\sqrt{1-u^2/c^2}$$

whence

$$\left.\begin{aligned} v &= \frac{u}{\sqrt{1-u^2/c^2}} \\ u &= \frac{v}{\sqrt{1+v^2/c^2}} \\ \sqrt{1+v^2/c^2} &= \frac{1}{\sqrt{1-u^2/c^2}}. \end{aligned}\right\} \quad (12)$$

Thus if v is the true velocity, u is not the velocity. The limitation of u to values below c that is a major point in Einstein's theory imposes no limitation on v; we see from equations (12) that as $u \to c$, $v \to \infty$. Also, if u is not the velocity, the acceleration is not u^2/ρ; taking it as v^2/ρ and using equations (12), the acceleration is found to be $u^2/[\rho(1-u^2/c^2)]$. From equation (8), with equations (12), we obtain

$$\frac{q_e E}{\sqrt{1-u^2/c^2}} = m_e \frac{u^2/\rho}{1-u^2/c^2} \quad (13)$$

agreeing with equation (10). On the right-hand side the acceleration is multiplied by the mass m_e; there is no 'relativistic' mass increase, which is an illusion generated by the Lorentz force law (which has not been established for appreciable values of u/c) and the belief that u is the velocity. But we see from the left-hand side of equation (13) that the electric force is not $q_e E$ as equation (2) states, but $q_e E/\sqrt{1-u^2/c^2}$. The information available to Lorentz was not good enough to enable him to arrive at this formula, but to take the force as $q_e E$ is permissible only for sufficiently small u (or v), and the assumption that this value applies for all values of u cannot be justified.

3.3 Motion of an Electron Parallel to an Electric Field

We imagine an electron moving along the x axis under the action of an electric field $E(x)$, directed along x, with a potential $V(x)$ which is zero at $x = 0$. The electron starts with zero velocity at $x = 0$ at time $t = 0$. After a time t it is at x, where the potential is V, the field is $E = -\mathrm{d}V/\mathrm{d}x$, and the velocity of the electron is v.

The force on the electron is then

$$\begin{aligned} F_E &= -q_e E\sqrt{1+v^2/c^2} \\ &= q_e \frac{\mathrm{d}V}{\mathrm{d}x}\sqrt{1+v^2/c^2} = m_e\,\mathrm{d}v/\mathrm{d}t. \end{aligned} \quad (14)$$

The work done on the electron as it moves from x to $x+\mathrm{d}x$ is

$$\mathrm{d}W = q_e \frac{\mathrm{d}V}{\mathrm{d}x}\mathrm{d}x\sqrt{1+v^2/c^2} = m_e \frac{\mathrm{d}v}{\mathrm{d}t}\mathrm{d}x. \quad (15)$$

Then

$$q_e \frac{\mathrm{d}V}{\mathrm{d}x}\mathrm{d}x = \frac{m_e\,\mathrm{d}v/\mathrm{d}t}{\sqrt{1+v^2/c^2}}\mathrm{d}x$$

whence

$$q_e \frac{\mathrm{d}V}{\mathrm{d}x}\frac{\mathrm{d}x}{\mathrm{d}t}\mathrm{d}t = \frac{m_e\,\mathrm{d}v/\mathrm{d}t}{\sqrt{1+v^2/c^2}}\frac{\mathrm{d}x}{\mathrm{d}t}\mathrm{d}t$$

i.e.

$$q_e \frac{\mathrm{d}V}{\mathrm{d}t}\mathrm{d}t = \frac{m_e\,\mathrm{d}v/\mathrm{d}t}{\sqrt{1+v^2/c^2}}v\,\mathrm{d}t.$$

Now integrate with respect to t.

$$q_e V = m_e c^2\sqrt{1+v^2/c^2} + \Gamma.$$

Γ is a constant of integration given by the fact that at $t = 0$ the electron is at $x = 0$ where $V = 0$ and its velocity v is 0. Hence $\Gamma = -m_e c^2$ and we finally obtain, for the relation between velocity and voltage drop,

$$q_e V = m_e c^2\{\sqrt{1+v^2/c^2} - 1\}. \quad (16)$$

For the energy of the electron, we return to equation (15) and integrate:

$$W = \int_0^x q_e \frac{\mathrm{d}V}{\mathrm{d}x}\sqrt{1+v^2/c^2}\,\mathrm{d}x.$$

The factor $\sqrt{1+v^2/c^2}$ can be expressed in terms of V by means of equation (16). Hence

$$W = q_e \int_0^V [1 + q_e V/m_e c^2]\,\mathrm{d}V,$$

whence

$$W = q_e V[1 + q_e V/2m_e c^2]. \quad (17)$$

Using equation (16), equation (17) yields

$$W = \tfrac{1}{2}m_e v^2. \quad (18)$$

Correspondingly, the Maxwell–Lorentz theory gives

$$W = q_e V = m_e c^2\left\{\frac{1}{\sqrt{1-u^2/c^2}} - 1\right\}. \quad (19)$$

Bearing in mind equations (12), we see that equation (16) agrees with the Lorentz formula for the values of $q_e V$, but that in the new theory this is not the energy of the electron, which is given by equation (18), agreeing with Newtonian mechanics. In view of this, it is seen that the practice of expressing energy in units of electron-volts is misleading.

3.4 The Magnetic Field

The basis of magnetism is Ampère's law; given this, the whole of magnetism, as known at present, can be developed. Thus to account for the phenomena of magnetism it suffices to derive Ampère's law from equation (6).

Consider an element δl_1 of conductor along which a current flows consisting of n_1 electrons per unit length moving at velocity v_1. These electrons are approximately balanced by n_1 fixed positive charges per unit length. Let there be another element δl_2 at a distance r from δl_1.

The positive charges in δl_1 exert a force F_1 on the positive charges in δl_2 and a force F_2 on the moving negative charges in δl_2. The negative charges in δl_1 exert forces F_3, F_4, on the positive and negative charges, respectively, in δl_2. Using equation (6) we obtain

$$\left.\begin{aligned}
&\text{(a)}\quad F_1 = \frac{n_1 q_e \delta l_1 n_2 q_e \delta l_2}{4\pi\varepsilon_0 r^2}\\
&\text{(b)}\quad F_2 = -\frac{n_1 q_e \delta l_1 n_2 q_e \delta l_2}{4\pi\varepsilon_0 r^2}\sqrt{1+v_2^2/c^2}\\
&\text{(c)}\quad F_3 = -\frac{n_1 q_e \delta l_1 n_2 q_e \delta l_2}{4\pi\varepsilon_0 r^2}\sqrt{1+v_1^2/c^2}\\
&\text{(d)}\quad F_4 = \frac{n_1 q_e \delta l_1 n_2 q_e \delta l_2}{4\pi\varepsilon_0 r^2}\sqrt{1+|\mathbf{v}_1-\mathbf{v}_2|^2/c^2}
\end{aligned}\right\} \quad (20)$$

and the net force of element 1 on element 2 is $F = F_1+F_2+F_3+F_4$. If θ is the angle between the directions of $\mathbf{v}_1$ and $\mathbf{v}_2$, the square of the relative velocity of the negative charges is

$$\begin{aligned}|v_1-v_2|^2 &= (v_2-v_1\cos\theta)^2+v_1^2\sin^2\theta\\ &= v_1^2+v_2^2-2v_1v_2\cos\theta.\end{aligned}$$

Making this substitution and expanding the square roots in equations (20) as far as second-order terms,† we obtain

$$F \simeq \frac{n_1 q_e \delta l_1 n_2 q_e \delta l_2}{4\pi\varepsilon_0 r^2}\left\{1-\left(1+\frac{1}{2}\frac{v_1^2}{c^2}\right)-\left(1+\frac{1}{2}\frac{v_2^2}{c^2}\right)+\left[1+\frac{1}{2}\frac{v_1^2}{c^2}+\frac{1}{2}\frac{v_2^2}{c^2}-\frac{v_1v_2}{c^2}\cos\theta\right]\right\},$$

i.e.

$$F \simeq -\frac{v_1v_2}{c^2}\cos\theta\,\frac{n_1 q_e \delta l_1 n_2 q_e \delta l_2}{4\pi\varepsilon_0 r^2}.$$

But $nq_e v$ is the current, i, so

$$F \simeq -\frac{i_1\delta l_1 i_2 \delta l_2 \cos\theta}{4\pi\varepsilon_0 c^2 r^2}.$$

We now define a new constant μ_0 by

$$c^2 = 1/\varepsilon_0\mu_0 \quad (21)$$

and obtain

$$F \simeq -\frac{\mu_0 i_1\delta l_1 i_2\delta l_2\cos\theta}{4\pi r^2} \quad (22)$$

which is Ampère's law.

If $\mathbf{r}$ is a unit vector from δl_1 along the direction towards δl_2, equation (22) can be written

$$\mathbf{F} = \frac{\mu_0\{i_2\mathrm{d}\mathbf{l}_2\times(i_1\,\mathrm{d}\mathbf{l}_1\times\mathbf{r})\}}{4\pi r^2} \quad (23)$$

and the magnetic field acting on $i_2\,\mathrm{d}\mathbf{l}_2$ is then

$$\mathrm{d}\mathbf{H}_1 = \frac{i_1\,\mathrm{d}\mathbf{l}_1\times\mathbf{r}}{4\pi r^2} \quad (24)$$

which is the Biot–Savart law. By integrating equation (24) over all current elements, the total magnetic field can be obtained.

3.5 Motion of a Charged Body in a Magnetic Field

For the force on a charged body—e.g. an electron—in a magnetic field, we consider first the force on it due to a current element. The total force due to a field will then be the sum of the elementary forces due to all the current elements in the system of coils and permanent magnets generating the field. For an electron moving in the neighbourhood of a current element, the element $i_2\delta l_2$ is replaced by a single charge $-q_e$ moving with velocity v_2. The forces F_2 and F_4 in equations (20) are then the forces on the charge, while F_1 and F_3 do not exist. The net force on the electron is then F_2+F_4, i.e.

$$\mathrm{d}F_M = \frac{n_1 q_e^2\delta l_1}{4\pi\varepsilon_0 r^2}\{-\sqrt{1+v_2^2/c^2}+\sqrt{1+|v_1-v_2|^2/c^2}\}. \quad (25)$$

Again v_1/c, the velocity of the electrons associated with current flow in the element $i_1\delta l_1$, is very small, so that second-order terms in v_1/c can be neglected. Expanding equation (25) to the first order in v_1/c we obtain

$$\mathrm{d}F_M = -\frac{n_1 q_e\delta l_1}{4\pi\varepsilon_0 r^2}v_1\cos\theta\,\frac{v_2/c^2}{\sqrt{1+v_2^2/c^2}}. \quad (26)$$

Putting $n_1q_ev_1\delta l_1 = i_1\delta l_1$, using equation (24), and writing $\mathbf{dB} = \mu_0\,\mathbf{dH}$, we obtain for a particle of charge $-q_e$ moving with velocity $\mathbf{v}$ in a field $\mathbf{dB}$

$$\mathbf{dF}_M = -\frac{q_e\mathbf{v}\times\mathbf{dB}}{\sqrt{1+v^2/c^2}} \quad (27)$$

or, summing over all current elements,

$$\mathbf{F}_M = -\frac{q_e\mathbf{v}\times\mathbf{B}}{\sqrt{1+v^2/c^2}}. \quad (28)$$

† To neglect higher-order terms in v_1/c and v_2/c is justified, since the velocity of electrons in current-carrying conductors is always very small compared with c.

It is understood that all the current elements involved are at rest with respect to one another. If not, extra velocity terms must be incorporated in equations (20).

An electron moving at right angles to a magnetic field **B** will follow a circular path of radius r. The force F_M on the electron will be balanced by the centrifugal force $m_e v^2/r$, where v^2/r is the acceleration. Hence

$$\frac{q_e vB}{\sqrt{1+v^2/c^2}} = \frac{m_e v^2}{r}, \tag{29}$$

whence

$$Br = \frac{m_e v}{q_e}\sqrt{1+v^2/c^2}. \tag{30}$$

According to the Maxwell–Lorentz theory, the force is

$$\mathbf{F}_M = -q_e \mathbf{u} \times \mathbf{B} \tag{31}$$

which, with equations (12), agrees with equation (28). The mass is $m_e/\sqrt{1-u^2/c^2}$ and the acceleration u^2/r. Hence

$$q_e uB = \frac{m_e u^2/r}{\sqrt{1-u^2/c^2}} \tag{32}$$

and

$$Br = \frac{m_e u/q_e}{\sqrt{1-u^2/c^2}}. \tag{33}$$

Equations (32) and (33) do not agree with equations (29) and (30) if u and v are related by equations (12). Discussion of this point is postponed to Section 4.

3.6 Maxwell's Equations and the Force Law

An electric field is defined by the force on a unit charge *at rest* (with respect to the electrodes generating the field) by putting $v = 0$ in equation (7). Alternatively, it may be defined as the force on a unit charge *at rest* (with respect to the charge q_1), in accordance with equation (6) when $v = 0$, summing over all other charges, of which q_1 is one, the charges all being at rest with respect to one another.

A magnetic field is defined by equation (24), summing over all current elements. Equation (24) has been developed from equations (20) and holds for small v_1/c. In practice v_1/c will be small, but this limitation means that a magnetic field is an artificial concept, not having the same reality as an electric field.

Subject to these limitations, Maxwell's equations can be developed from Coulomb's law and Ampère's law as in classical electromagnetism. What the new theory does is give criteria for the validity of Maxwell's equations; they are valid only for small velocities, which is to say small currents and small rates of change of field, magnetic or electric.

The limitation to small rates of change (small $\partial/\partial t$) suggests that the derivation of a wave equation for electromagnetic waves from Maxwell's equations may be valid only for sufficiently low frequencies. Which may have a bearing on the wavy character of electromagnetic radiation at microwave frequencies and below, and its particle-like character at optical frequencies and above. In this respect it may also be significant that in the lower frequency range the radiation is generated by the methods of electrical circuitry, while in the higher frequency range it is generated by quantum processes.

Since Coulomb's law is a special case of equation (7) and Ampère's law is derivable from that equation, it follows that Maxwell's equations amount to no more than a development of equation (7), valid only for small v. Also, the law of force on a moving charge in a magnetic field is already contained within equation (7), which replaces the Lorentz force formula. Thus equation (7) replaces Maxwell's equations and the Lorentz force formula; it not only yields equations (1) and (20), but enables us to appreciate the range of their validity.

3.7 Time Dilation

According to the Lorentz theory of electrons and to Einstein's special theory of relativity, if a clock moves with velocity u relative to an observer, it appears to that observer to be slowed by a factor $\sqrt{1-u^2/c^2}$ with respect to its rate when at rest with respect to the observer. This, it is claimed, is exemplified by the behaviour of fast mesons in storage rings. A μ meson at rest has a decay time τ of the order of 2 μs. In a storage ring, where, it is believed, the meson travels at a velocity u which is close to c, the decay time, as observed by an observer at rest in the laboratory, is many times τ. If l is the length of the track round the ring, and if the meson makes n circuits before decaying, its decay time is

$$\tau' = nl/u \tag{34}$$

and this is observed to be many times τ. The orthodox theory requires

$$\tau' = \tau/\sqrt{1-u^2/c^2} \tag{35}$$

and so

$$\tau = \frac{nl}{u}\sqrt{1-u^2/c^2}. \tag{36}$$

τ' is not observed; what is observed is a value of n very much greater than would be expected if the decay time remained equal to τ, so that n would be equal to $u\tau/l$. The apparently dilated decay time τ' is obtained by observing n; then, knowing τ and l, u is calculated from equation (36), and then equation (35) gives τ'.

Newtonian theory would require, for mesons travelling round the ring with velocity v,

$$\tau = nl/v \tag{37}$$

in place of equation (36). These equations agree if equations (12) hold good. There is no time dilation; the mesons make many more circuits of the ring than would be expected because they are travelling much faster than is generally believed—v is much greater than u, or much greater than c since u is close to c.

4 Discussion

If we use equations (12), we find that most of the equations in u in Section 3 agree with the corresponding equations in v. The only pair which do not agree are equations (29) and (32). If we look more closely at the equations which agree, some further understanding can be obtained which will help us to see what is wrong in the case of (29) and (32).

The straightforward comparison of u and v in equations (16) and (19), (28) and (31), and (36) and (37), as well as the derivation of Ampère's law and of the magnetic force formula (eqn. (28)), reinforce our confidence in the rightness of equations (12) and of the identification of v, not u, with the true velocity.

By writing equation (10) in the form of equation (13) we are able to determine that the acceleration v^2/ρ of a particle in a curved path is given in terms of u as $(u^2/\rho)/(1+u^2/c^2)$. Then, in equation (32), the right-hand side is not mass times acceleration; to make the right-hand side equal to mass times acceleration, we should, perhaps, write

$$\frac{quB}{\sqrt{1-u^2/c^2}} = \frac{m_e u^2/r}{1-u^2/c^2} \qquad (38)$$

but then the left-hand side is not equal to the Lorentz force nor to the value given by equation (28). We can get a formula in which the force is expressed as mass times acceleration by taking the force as in equation (31) and the mass times acceleration as in equation (38), thus

$$quB = \frac{m_e u^2/r}{1-u^2/c^2} \qquad (39)$$

which now agrees with equation (29) but whose only justification is that it does so agree.

This disagreement is disturbing, but the fault lies with equation (32), not with the new theory. Equation (32) is inconsistent with all the other equations in terms of u, i.e. with equations (10), (19), and (36). If $(u^2/\rho)/(1-u^2/c^2)$ is the acceleration in the electric case, $(u^2/r)/\sqrt{1-u^2/c^2}$ cannot be the acceleration in the magnetic case. But if the acceleration is written as on the right-hand side of equation (39), then the left-hand side of equation (39) is not the force; the new theory and the Lorentz theory agree on this point.

Nowadays a great many experiments are performed using charged particles which are accelerated to high velocities. In linear accelerators particles are constrained to have a velocity close to c, but their energy is not measured so that the claim that it is so many electron volts is unfounded. To conclude anything useful, at least two measurements should be made, on the same particles in a single experiment, out of the following: the curvature of the path in a transverse electric field; the longitudinal voltage drop in an accelerator using d.c., not r.f. which offers the chance of error due to uncertainties in phase; the path curvature in a transverse magnetic field; the time of flight over a known distance. Almost always only one of these quantities is observed. Usually the curvature of a track in a magnetic field is observed, and the velocity inferred from equation (32). Scientists are so sure that equation (32) is correct that they think it unnecessary to check it by simultaneously measuring the velocity directly. To be sure, to do so would give a great deal of trouble. So does being mistaken about one's assumptions.

There are a few rare experiments in which useful information is apparently obtained. Champion[12] measured the curvature of β particle tracks in a magnetic field, and then the β particles struck stationary electrons and were scattered, the angle between the paths of the scattered β particles and the recoil electrons being observed. The 'relativistic' relation between the particle velocity and the angle of separation was confirmed. Rogers, McReynolds, and Rogers[13] performed two separate experiments using the β particles from a radium source, characterized by three sharply defined lines in the energy spectrum. In one experiment $E\rho$ was measured, in the other Br. From equations (11) and (33) we have

$$\frac{E\rho}{Br} = u \qquad (40)$$

and

$$\frac{B^2r^2}{E\rho} = \frac{m_e/q_e}{\sqrt{1-u^2/c^2}}. \qquad (41)$$

Rogers *et al.* determined u by equation (40) and then checked (41), finding it to be satisfied by the value of u determined from (40). But the quantity called u in equation (31) is not the same as the u of equation (11), and consequently the u of equation (40) is equal to neither of them. It is therefore difficult to understand how equations (40) and (41) can be compatible.

In view of the doubts raised by the new theory and the discussion of the Lorentz theory given here, it is important that these experiments should be repeated, and that others should be performed. In particular, reports of experiments should give the observations made as well as the conclusions drawn from them. Velocities should only be mentioned if time-of-flight over a known distance has been directly observed. Energy should only be mentioned if it has been measured calorimetrically. The usual practice of measuring track curvature in a magnetic field and calling it velocity, or observing voltage drop (if, indeed, it is actually observed, as against being calculated or assumed) and calling it energy, are rituals, not scientific procedures, and the result is that modern physical theory is in a state of chaos which will take decades to put right, even if the will is there to do so.

5 References

1 Waldron, R. A., 'The Wave and Ballistic Theories of Light—a Critical Review' (Frederick Muller, London, 1977).

2 Waldron, R. A., Letters in *Electronics and Power*, **10**, p. 92 and p. 168, 1964.

3 Dingle, H., 'Science at the Crossroads' (Martin Brian and O'Keeffe, London, 1972).
4 Parish, L., 'The Logical Fallacies of Einsteinian Relativity' (Luton, 1977).
5 Palacios, J., 'Relatividad—una Nueva Teoria' (Espasa-Calpe, Madrid, 1960).
6 Dingle, H., (*op. posth.*), 'The "twins" paradox of relativity', *Wireless World*, **86**, pp. 54–6, October 1980.
7 Nedved, R., 'Classical Theory of Relativity', *Naucna Misao (Zagreb)*, Nos. 14/15, 1978/9.
8 McCausland, I., 'Science on the defensive', *Can. Elect. Engrs J.*, **5**, no. 2, pp. 3–4, 1980.
9 Burniston Brown, G., 'What is wrong with relativity?', *Bull. Inst. Phys. & Phys. Soc.*, **18**, pp. 71–7, March 1967.
10 Burniston Brown, G., 'Einstein versus thought-experiment', *Am. J. Phys.*, **44**, pp. 801–2, 1976.
11 Trouton, F. T., and Noble, H. R., 'The forces acting on a charged condenser moving through space', *Proc. Roy. Soc., A* **72**, pp. 132–3, 1903.
12 Champion, F. C., 'On some close collisions of fast β particles with electrons, photographed by the expansion method', *Proc. Roy. Soc.*, A **136**, pp. 630–7, 1932.
13 Rogers, M. M., McReynolds, A. W., and Rogers, F. T., 'A determination of the masses and velocities of three radium β particles—the relativistic mass of the electron', *Phys. Rev.*, **57**, pp. 379–83, 1940.

Manuscript received by the Institution on 31st March 1981.
(Paper No. 2114)

Book Review

James Clerk Maxwell. A Biography

Ivan Tolstoy

Canongate, Edinburgh, 1981. 13.5 × 22 cm, 184 pages, £9.95*

In this 150th anniversary year of the birth of James Clerk Maxwell the appearance of a new biography is both appropriate and welcome. Welcome because the author has followed an approach to his subject distinctly different from previous biographers, to whom he gives generous acknowledgment. His modestly phrased aim is that it 'pretends neither to be a definitive biography, nor a work of historical scholarship. It is, rather, a book for the lay reader'. But it is not merely that, because in stressing the whole Maxwell Professor Tolstoy has, to use the modern metaphor, shown what made him 'tick', as far as is possible in writing about an essentially 'private man'. The influences of his boyhood, with its somewhat rugged upbringing, his early blossoming into a theoretical physicist who was nevertheless an adept experimenter, his university posts, his marriage and his deep religious convictions, are sensitively and perceptively woven together. The reader (or this reviewer at least) thus feels that a revealing window has been opened on the real Maxwell.

Professor Tolstoy, himself a distinguished applied mathematician and goephysicist who was trained and has worked in both Europe and the United States, deals with university life of the middle of the last century in an absorbing way. He concludes that while Maxwell seems not to have been a good lecturer—and indeed was found frequently difficult to follow in conversation by all except his peers—his powers of exposition on paper were of the highest clarity possible in abstruse subjects. He appears to have been rather at sea in the politics of academe but on the credit side his planning and supervision of the Cavendish Laboratory is impressive.

But this is not just another biography of an 'eminent Victorian'—how could it be when so much needs be interpreted of the towering scientific contributions of Maxwell, albeit with the lay reader in mind? If the physicist wishes to go deeply into these matters, the annotated bibliography indicates the recognized sources. However, for the engineer, the 'mix' is about right.

Throughout the book one is amazed at the range of subjects on which Maxwell worked and published significant results. Thus we learn of the early work in geometry, the prize for explaining the stability of the rings of Saturn, the dynamical theory of gases, experiments on colour vision, and then of the gradual evolution of the electromagnetic theory of light and its mathematical development in the famous equations. Later on came the 'Treatise on Electricity and Magnetism', the theory of governors, more work on gases, and a paper on Boltzmann's theorem which was significant in founding statistical mechanisms.

Though not in the generally accepted sense a polymath, the descriptions, 'a physicist's physicist' given by Professor Tolstoy and 'the complete physicist' taken by Professor R. V. Jones for the title of his Memorial Lecture, both sum up a scientist of the first rank whose three decades of productive work changed the face of natural philosophy.

The book is readable and well written: the production is dignified and the few printer's errors noted will not mislead (nor, probably, will the placing of Keswick first in the Lake District, and in the next paragraph in the Peak District!). It is warmly recommended to the dedicated readers of biographies and it will be found revealing and rewarding reading for those who seldom feel inclined to spare time for other than technical matters. The coincidence of the publication of Professor Jones's Memorial Lecture in this issue of the Journal provides a shorter but complementary account which may well lead the latter category of radio and electronic engineers to want to know more about the scientist to whose work their profession owes so much.

F. W. Sharp

*The publishers are offering this book to members of the Institution at the special price of £7.95 post free until 31st March 1982. Send cheques or postal orders to Canongate Publishing, 17 Jeffrey Street, Edinburgh EH1 1DR marking 'IERE Offer' on the envelope.

UDC 621.391.812.7
Indexing Terms: Radio systems, diversity, Computational methods

Optimum diversity separation for over-sea line-of-sight radio links

A. J. HENK, C.Eng., M.I.E.R.E.*

SUMMARY

Given the requirement for a line-of-sight link operating over a sea path and having established the clearance at the horizon over the appropriate range of k-values using standard procedure, a method is derived for determining the optimum spacing for worst case conditions between receiving antennas where dual space diversity is employed to reduce 2-path fading. The means of calculating the minimum received signal, i.e. the deepest fade possible with destructive interference in the 2-path mode, is also included; as a check on performance if the optimum spacing cannot be realized or where the optimum value is used and the worst case conditions assumed in this determination do not apply in practice.

Programs for these calculations are included which run on TI58 or TI59 pocket calculators. A third program for the calculation of Fresnel zone numbers (cubic equation solution) is included. The method is valid for any frequency band in which line-of-sight criteria apply and 2-path propagation occurs.

Although consideration of the probability of occurrence of any particular propagation mode lies outside the scope of this paper, it is found that the correct 2-path diversity spacing also provides substantial protection in other adverse propagation circumstances.

* Brown & Root (U.K.) Ltd., Brown & Root House, 125 High Street, Colliers Wood, London SW19 2JR.

Symbols and Abbreviations

d	great circle distance
f	frequency
h_a	lower antenna elevation
h_b	upper antenna elevation
k	effective Earth's radius factor (i.e. effective Earth's radius = kr)
M	profile of the atmospheric refractive index against height
n	Fresnel zone number
S	spacing between upper and lower receive antennas ($= h_{2b} - h_{2a}$)
t	tidal height: in this connection it refers to the total height of the sea above l.a.t. including waves, swell, etc.
f.s.l.	free-space loss
f.s.r.	received signal strength under conditions of free-space propagation
l.a.t.	lowest astronomical tide
l.o.s.	line-of-sight
α_c	proportion of Fresnel zone spanned by abscissa values for the coarse zone structure
α_f	as above, but for the fine zone structure

Natural variables: Environmental variables not under the control of the designer and which affect system performance. In this case they are k and t.

1 Introduction

Oil comes out of the ground in strange places: so much so that the building of a small town supported on huge stilts more than a hundred miles out into the ocean is now accepted as a commonplace procedure for winning supplies of energy from ten thousand feet below almost 100 fathoms of inhospitable North Sea. Not surprisingly a whole technology has grown up around offshore platform design: the provision of an effective telecommunications network throughout the North Sea forms an essential part of this technology. In the UK sector British Telecom is responsible for the provision of the onshore facilities and for the overall network philosophy, which includes tropospheric scatter circuits for long haul (trans-horizon) routes and u.h.f. line-of-sight for local working, operating in the L-band.[1-4] Of these two techniques it is the line-of-sight network, whose systems greatly out-number the tropo installations, which concerns us and in particular that aspect of l.o.s. operation peculiar to offshore work where the path is entirely over water.

Fortunately there is now available, in the form of the programmable pocket calculator, an impressively powerful design tool which permits hitherto impractical analyses and calculations to be performed in a matter of minutes. This enables the systems engineer to take into account, for worst-case conditions, that particular problem which affects microwave links over water: surface reflections.

Classic papers by Vigants[5] and later Lewin[6] address this problem and serve to illustrate the difficulties inherent in defining an optimum design, particularly in the selection of the spacing for the vertically separated receiving antennas. Historically there has been a tendency for designers to use methods normally applicable to overland paths and based on stochastic criteria, with follow-up examination of the effects of reflection at one or two spot combinations of tide (t) and effective Earth's radius (kr) values to check that both antennas do not fade simultaneously. This involves a measure of trial and error which increases in complexity as attempts are made to improve the system by checking for possible destructive interference over a wider range of natural values. Whereas the method to be described does not completely eliminate the need for trial and error, this process is carried out in the calculator by the technique known (euphemistically) as 'iterative programming'.

The first steps in deriving this approach are an examination of the propagation path, in particular the effects of sea-surface reflections, and the definition of some sort of optimum criterion to be approached as part of the design.

2 Applicability and Limitations

Electromagnetic radiation from a microwave line-of-sight antenna can reach a receiving antenna by various routes. The most significant route (by definition) in an l.o.s. system is direct: see Fig. 1. This is the free-space ray whose attenuation due to propagation (the free-space loss) is given by:[7]

$$\text{f.s.l.} = 32{\cdot}4 + 20\log d + 20\log f \quad \text{dB}$$

where d is the great circle distance in km ($d = d_1 + d_2$) and f is the frequency in MHz.

From the free-space loss the signal present at the receiver input terminals, under free-space conditions, (f.s.r.) can be calculated by summing the feeder and associated losses together with the antenna gains, expressed in dB, and the transmitter output power expressed as dBm or dBW.

The second route is by reflection from the sea surface, by which signals can arrive at the receiving antenna with any value of phase relative to the direct ray, and any signal strength from f.s.r. downwards. The worst case is where the difference in path length between the direct

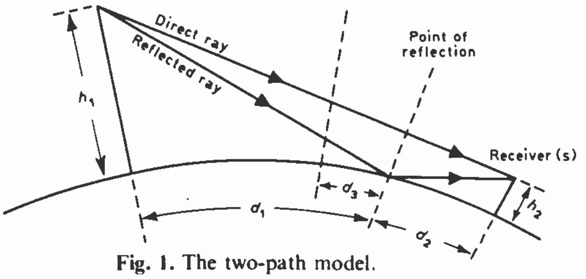


Fig. 1. The two-path model.

and reflected rays and the phase change on reflection result in a phase difference (ϕ) of π radians at the receiving antenna, with a reflected signal strength of f.s.r., leading to cancellation. The value of ϕ can be found by calculation; the value of reflected signal strength is dependent upon the state of the sea surface, which partly determines the coefficient of reflection. In the worst case this coefficient approaches -1, and this is the value assumed during the following calculations. Also assumed, as a worst-case possibility, is a combination of fading pattern phases (the significance of which will become clear in Section 5) resulting in equally deep fades at the extremes of the ranges of the natural variables.

Some links, however, will be found to have a geometry for which the resultant signal level remains at or above f.s.r. when the correct diversity design is applied. The program test run data in Appendix 3 are for such a link. In a case like this the lowest resultant signal corresponds to f.s.r. and occurs when there is no reflection because the reflected signal reinforces the direct signal in all other circumstances considered. The lowest signal as calculated by the programs, being based on total reflection, should be interpreted with this in mind: the effect is not usually significant in practice and only occurs when the lowest calculated signal exceeds f.s.r.

There are further mechanisms for the propagation of energy between the antennas which depend upon anomalies in the rate of change of atmospheric refractive index with height (profile M: ref. 7). Adverse M-profiles can deflect the direct ray and/or cause reflections from within the atmosphere. It is these M-related mechanisms which may lead to the requirement on overland paths for diversity protection the extent of which is based on the decorrelation between spaced antennas. This is determined using stochastic techniques.[5-9]

On sea-links, since the mechanism for potentially strong surface reflections is present, the two-path model represents the dominant mode of propagation, Laine[10] giving values worse than $-0{\cdot}9$ for the reflection coefficient of smooth sea and Vigants[8] stating discouragingly: 'Over-water paths, undesirable because of reflections, are sometimes unavoidable.'

In view of the relatively large number of radio links operating in the North Sea it is both surprising and disappointing that more propagation data are not available. The importance of the two-path situation is further exemplified by Inoue and Akiyama[11] in measurements made on oversea paths near Tokyo in 1970 and 1971. Tests over one four-week period at 6·72 GHz identified significant two-path fading as being present for more than 83% of the time and less severe two-path fading (fade depths of less than 5 dB) for 16%. Over two periods in 1970 and 1971 the cumulative distribution shows that a reflection coefficient of $-0{\cdot}8$ or worse was present for 30–40% of the time. Comparison with results of similar measurements at 20 GHz show a marked dependence upon wavelength, strongly suggesting that the probability of near-specular reflection at 2 GHz is even greater than at 6·72 GHz. The protection against two-path propagation provided by the following method is also consistent with classical protection against *M*-related fading, particularly on longer sea-paths where substantial decorrelation results. On short hops the incidence of *M*-related effects will be lower, but severe fading is still experienced: the two-path model still applies and full protection against two-path fading is maintained.

The natural variables which need to be taken into account in the planning of sea-links are the effective Earth's radius factor, k, and tide, t. In this connection tide includes swell, wave height and other changes in local sea-level. The problems given in the appendices allow for k-values from 0·8 to 10^6 and for t-values of zero to 20 metres, so that full protection exists throughout theses ranges. The values can be changed if required simply by substituting the requisite numbers for those in the corresponding program locations: for example the program utilizes $k = 0{\cdot}8$ as a minimum value in accordance with the British Telecom planning criteria for local sea links,[4] but many planners prefer to use 0·6. The tide range of 0–20 metres is more than adequate for North Sea use and is related to lowest astronomical tide (l.a.t.) rather than mean sea level (m.s.l.). Reference to m.s.l. can be made by inserting the appropriate values into the programs (see Appendices). Note that the same reference datum must be used for antenna heights as for the tide range. Negative t values are permissible.

Line-of-sight links operate by definition above grazing incidence, i.e. some positive clearance at the horizon. The minimum clearance chosen depends upon the approach of the system designer and will not be considered here, but it is of course necessary for the lowest antenna height to be adequate to ensure an above grazing path at the lowest k and the highest t values before l.o.s. criteria can be applied. (Some designers will accept a limit grazing path if vertical polarization can be used.)

3 Fading Zone Patterns: The Cubic Equation

For a given transmitting antenna height, the difference in path length between the direct and reflected rays at the receiving antenna varies with receiving antenna elevation, assuming that k and t are steady. As the receiving antenna elevation increases through points at which the phases differ by multiples of 2π radians, reflection (assuming the worst case coefficient of -1) results in equal strength antiphase signals at the receiver, and a resultant field of zero. At odd multiples of π the inverse occurs and the reinforcing signals add together to produce a resultant which is 6 dB above f.s.r.[7,12] This is shown in Fig. 2 for a linear y-axis scale. Reinforcement

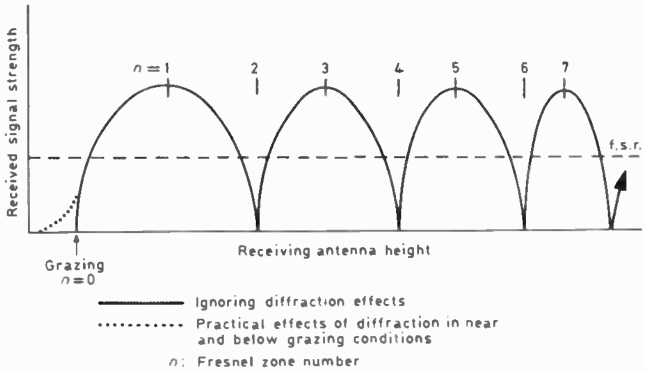


Fig. 2. Signal strength vs receiving antenna height.

at the receiver is maximum at odd Fresnel zone numbers and fades are deepest at even zones. It is anticipated that the reader will be familiar with the Fresnel zone structure, but a full treatment can be found in Bullington, CCITT, etc.[7,10,12] Although it is possible to calculate the antenna heights at which the various features of the zone structure are to be found, it is necessary to know, or at least to assume, values for k and t. Since each of these can vary sufficiently to affect not only the absolute position of the pattern in the vertical plane, but also the spacing between the adjacent peaks and troughs, the practical situation usually results in an uneven pattern running vertically across the antenna giving an unacceptable fading performance. In other words, wherever the antenna is located it will be wrong at least for some of the time. Nevertheless, in order to design a suitably protected system a fairly precise knowledge of the behaviour of the zone pattern is required.

As long ago as 1946 Millington[13] published an analysis of the path geometry which included a graphical means for the determination of the path difference between the direct and reflected rays from which a spot value performance can be readily deduced, but which rapidly becomes tedious if a range of k values is to be considered. To solve the problem without recourse to graphical methods requires the solution of a cubic equation which, to execute manually, is no less cumbersome. By the use of a calculator program, however, the solution becomes simple and rapid and the sub-routine for this purpose in the programs in the Appendices runs many times during the execution of the main program.

The version of the cubic equation used in these programs is that described by CCITT[7] and is solved as follows:

(1) Evaluate the parameters τ and r, where

$$\tau = \frac{d^2}{12} + 8{\cdot}5\,\frac{k}{4}\,(h_1 + h_2),$$

$$r = 6{\cdot}37\,\frac{k}{4}\,d(h_2 - h_1).$$

(2) Calculate d_3, the distance from the point of

reflection to the mid-path point, from

$$d_3 = 2\sqrt{\tau}\cos\left(\frac{\psi}{3} + 240°\right)$$

where

$$\psi = \cos^{-1}\frac{r}{\tau\sqrt{\tau}}.$$

(3) It is now a simple matter to calculate d_1 and d_2:

$$d_1 = \frac{d}{2} + d_3, \quad \text{and} \quad d_2 = d - d_1.$$

(4) Proceed to calculate the primed heights h'_1 and h'_2:

$$h'_1 = h_1 - \frac{4}{51}\times\frac{d_1^2}{k} \quad \text{and} \quad h'_2 = h_2 - \frac{4}{51}\times\frac{d_2^2}{k}.$$

(5) The Fresnel zone number can now be derived:

$$n = \frac{fh'_1 h'_2}{75d}.$$

Symbols in d and h are as in Fig. 1.

4 Diversity: Purposes and Features

The practice of adding a second receiving antenna separated vertically from the first gives access to a different part of the zone pattern, the intention being that the spacing is chosen such that only one antenna will fade at any one time. Lewin[6] describes a means of combining the antenna signals at carrier frequency using a variable phase-shifter controlled by a feedback signal from a single receiver. In a North Sea platform link, where a high order of reliability is required (99·99% time availability) it is standard practice to utilize two separate receivers whose baseband outputs are combined. In this case a receiver (or antenna) failure does not lead to complete loss of traffic although the system is now vulnerable to multipath fading.

Two methods of combining are in common use: optimal ratio combining and switched combining. In the former case the two receiver basebands are added in a linear mixing system in which the proportion of each receiver signal is controlled in accordance with the signal strength at the receiver concerned to keep the combined baseband level constant. The noise contributions from the receivers add incoherently whilst the traffic adds coherently and precautions are taken with the design of i.f. and a.g.c. circuits to ensure that the noise contribution from each receiver does not rise disproportionately as the signal fades. In this way the combined signal-to-noise ratio is always better than that of each receiver separately, particularly when the signal strengths are equal at each receiver. Should the noise factors of the receivers be unequal the system noise performance will vary smoothly between the two values dependent upon the relative signals, with best combined noise factor when the signal at the more noisy receiver is weaker by a small margin.

Switched combining selects one receiver at a time to feed the baseband output terminals in such a way that the receiver with the stronger signal is chosen. Although the switched combining receiver does not display the slightly improved noise performance which results from optimal ratio combining, it is a very effective technique as switching can be made fast and unobtrusive and it is simple to set up, not having any tracking problems. Switching occurs when the received signals are approximately equal so that sudden changes in background noise are not apparent in normal operation: however, the switching characteristics have to be controlled to a nicety if data are being transmitted over the link. The simplicity of switched diversity operation lends itself to illustration in the next Section in which the relationship between the diversity antenna arrangement and the Fresnel zone structure will be examined. It will be clear that, although the following discussion assumes switched diversity and the signal strengths calculated relative to f.s.r. are based upon this assumption, the optimum configuration of the diversity antennas given by the program applies equally to optimal ratio combining.

As a point of interest Sharman described in 1973[14] a diversity system in which typically four separate diversity signals whose phases are uncorrelated are combined, at the intermediate frequency of 70 MHz by an ingenious phase correction circuit, prior to demodulation. This is used in troposcatter reception and, so far as the author is aware, has little application in l.o.s. systems. Its implementation is generally as an optimal ratio combiner, and finds application in systems such as troposcatter in which operation close to the f.m. threshold is often encountered and where it effectively extends the range of operation possible by improving the carrier-to-noise ratio at the discriminator.

5 Optimum Diversity Separation

As an aid to identifying optimum diversity performance from which an optimum separation can be derived, it is helpful to consider a little more closely the way in which diversity improves the fading performance of an over-water l.o.s. link. A graphical approach has been found helpful as it readily allows for the discontinuity caused by switching from one antenna to the other whilst permitting the significance of this discontinuity to be observed. It must be remembered that, with given antenna heights, the natural variables not only cause the pattern of peaks and nulls to be swept across the antennas; they also cause changes in zone spacing to occur.

Furthermore, the zone spacing changes with receive antenna height (closer spacing at higher elevations) although at the larger zone numbers this effect becomes small. In Fig. 3(a) a signal strength profile is shown for the two-path condition. An antenna mounted at a height

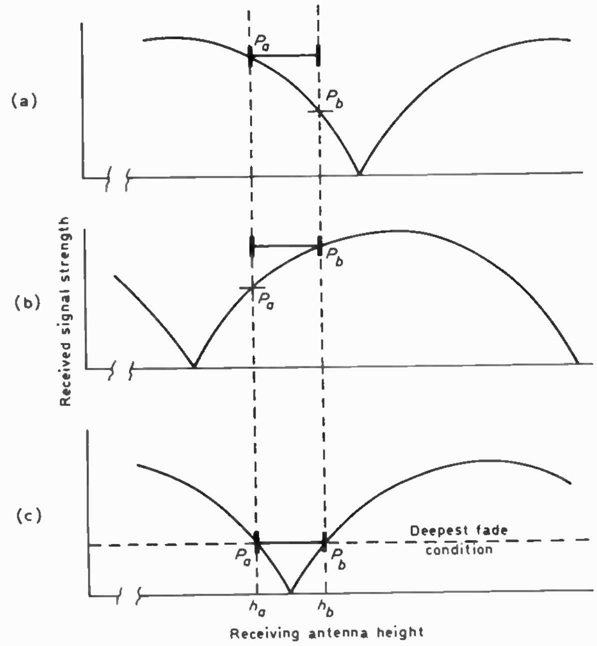


Fig. 3. Mechanism of diversity operation: coarse pattern.

of h_a will receive a signal whose value is given by P_a on the curve. A second antenna at height h_b will result in a signal corresponding to P_b, which is less than P_a. If these antennas form a diversity system the stronger signal will be selected, i.e. P_a. In Fig. 3(b) the fading pattern has moved so that the signal at the h_b antenna is now greater than at h_a, so that the upper is now selected. Since the zone spacing is significantly greater than the antenna separation this can be considered as a coarse pattern condition: a corresponding fine pattern condition can be identified in which the zone spacing approaches the separation. This is examined below.

If we now apply a simple construction and draw a short line of length $(h_b - h_a)$ parallel to the X-axis, we can allow this to ride vertically as the pattern beneath moves so that at least one end is always in contact with the curve. Thus if, in Fig. 3(b), the fading pattern were to move to the right in order to bring P_a into the trough, P_b would still contribute a useful signal. It is apparent now from Fig. 3(c) that the deepest fade encountered by the system occurs when the line $(h_b - h_a)$ straddles the null: it is unimportant which signal is selected as the two are equal. Clearly whichever way the pattern moves from this position, one of the antennas will receive a stronger signal. As the antenna spacing is reduced, i.e. as the line $(h_b - h_a)$ becomes shorter, the signal strength in this worst case becomes lower as the construction line sits further down the curves until, at zero separation, there is no protection and both antennas are in the null.

Let us now consider the situation in which the zone spacing is only slightly greater than the antenna separation as shown in Fig. 4. Here the deepest fade occurs when the antenna system straddles not a null but a peak. A small increase in antenna spacing now results in the two antennas entering adjacent nulls and once again there is no protection. Once this condition is reached the capability of the system has been exceeded and we can therefore identify the requirement that the antenna spacing must always be less than the finest zone pattern.

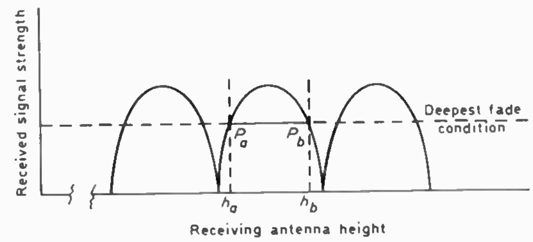


Fig. 4. Mechanism of diversity operation: fine pattern.

As the spacing $(h_b - h_a)$ increases from zero the minimum received diversity-protected signal also increases from zero, but at a more rapid rate for the finer pattern due to the greater steepness of the curves. When the 'over the top' condition of Fig. 4 is approached for the fine zone structure the minimum received signal starts to fall although the corresponding value for the coarse condition is still rising. A situation now exists in which any improvement in one case is made at the expense of the other. It is now possible to define an optimum condition for a certain range of zone pattern spacings, and this is that the spacing must be such that the minimum diversity-protected signal (i.e. deepest fade) in the coarse zone case equals that in the fine zone case.

An excellent set of curves has been published by Lewin as Figs. 2 and 3 of his paper (ref. 6, p. 298) in which the

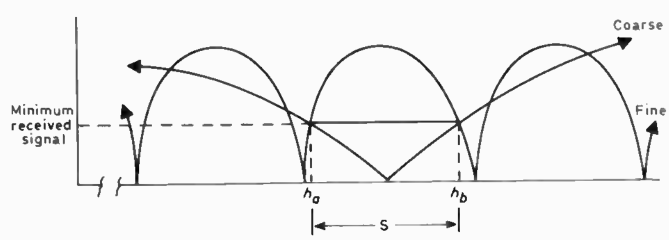


Fig. 5. Optimum diversity spacing.

received signal strengths for two diversity antennas are plotted against C $(= k^{-1})$. Although not a strict parallel to the case under discussion, the changes in phase and spacing of the fading patterns are clearly illustrated.

The diagram in Fig. 5 shows the fine and coarse situations superimposed with a phase relationship suitable for illustrating the condition of equal coarse and fine zone fades. It should be appreciated that the relative positions of the coarse and fine zones have been chosen to facilitate analysis and that this graphical construction does not necessarily represent the actual conditions. (It differs in two respects: coarse and fine patterns are not, of course, present simultaneously; also the phasing as drawn, while perfectly possible, is improbable.) What this construction does illustrate is the way in which any change from the spacing in the Figure makes one or other of the cases (coarse or fine) worse. Reducing spacing allows the construction line $(h_b - h_a)$ to ride further down into the trough on the coarse pattern, increasing it permits adjacent troughs on the fine pattern more nearly to be approached simultaneously.

We can now see clearly how protection afforded by the diversity principle is restricted by the effects of the natural variables as reflected by their influence on the extreme values of the Fresnel zone spacing which leads to the need to identify the range of zone spacings over which protection is required. Running a few examples in the cubic equation establishes the following:

(a) Increasing k increases the Fresnel zone number, other parameters being constant.

(b) Increasing t decreases the Fresnel zone number, other parameters being constant.

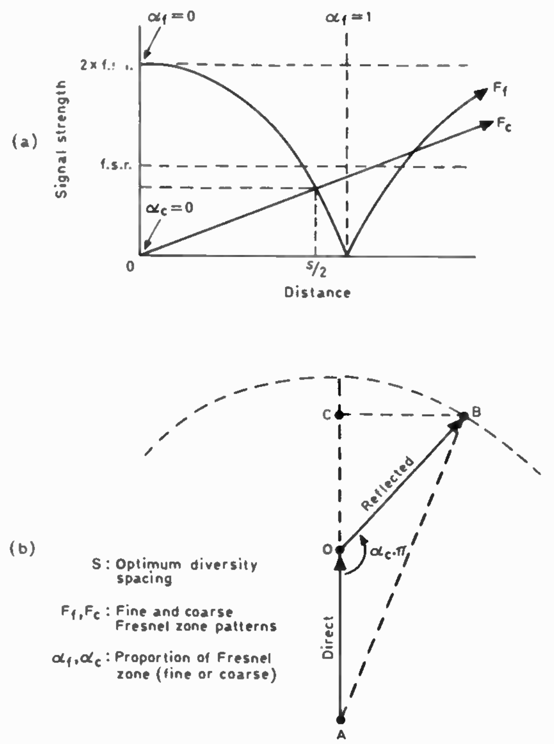


Fig. 6. Graphical construction for S.

By choosing the highest k value of interest and the lowest tide and applying these values simultaneously the finest zone structure can be established. Similarly, the minimum k and maximum t result in the coarsest case. From these extreme conditions a diversity spacing can be calculated which will give equal protection at each extreme and result in a superior performance at all intermediate values.

If the diagram in Fig. 5 is split along its axis of symmetry and redrawn as in Fig. 6(a) the spacing S can be calculated. In this diagram, α_f represents the proportion of the Fresnel zone being spanned by distances along the x axis for the fine pattern, similarly α_c applies to the coarse pattern. These proportions should ideally be expressed in terms of Fresnel zone numbers rather than linearly, as the zone spacing varies with distance. However, since we are now examining only a proportion of the total height encompassed by the calculations for n, and as it is the upper part of that height range which is being considered, we can introduce considerable simplification into the forthcoming manipulations of n and α by assuming a linear relationship between the two. This works well in practice because the effects of non-linearity become smaller at high zone numbers and the sensitivity of the minimum signal level to S is least for the coarsest pattern. In the fine pattern case, linearity is most closely approached where the spacing is most critical.

It is important to note that α_c and α_f are not the same as the zone numbers. but are related to the origin of the figure which has been chosen to facilitate the calculation. At the antenna heights at which the calculations are carried out, the zone number n can have any value and will be higher for the fine pattern than for the coarse. It must also be borne in mind that the phases of the fine and coarse patterns in Figs. 5 and 6 have also been chosen to facilitate worst-case calculation and in practice the patterns can have any relationship. The value of S, however, remains valid whatever this relationship may be in real life.

The phasors in Fig. 6(b) relate to the direct and reflected rays of the coarse pattern which cancel when $\alpha_c . \pi = 0$ and reinforce when $\alpha_c . \pi = \pi$, i.e. $\alpha_c = 0$ and 1 respectively.

Since the diagrams are normalized to f.s.r., phasors OA and OB are unity, so we have

$$\angle \text{BOC} = \arc\cos \text{OC} = \arc\sin \text{BC}$$

leading to an expression for the resultant normalized received signal strength AB

$$\text{AB} = \sqrt{(1+\cos(\pi - \pi\alpha_c))^2 + (\sin(\pi - \pi\alpha_c))^2}$$

from which

$$\text{AB} = \sqrt{2\cos(\pi - \pi\alpha_c) + 2}. \qquad (1)$$

The procedure for solving the fine structure is similar except for the phase shift of π called for by the graphical construction so that, by subtracting π from $(1-\alpha_c)\pi$, we can write

$$\text{AB} = \sqrt{2\cos(\pi\alpha_f) + 2}. \qquad (2)$$

Reference to Fig. 6(a) reveals that, as $S/2$ increases from zero and the value of received signal rises from zero on the coarse pattern as it falls on the fine pattern, the correct point is the first encounter with the equal signal condition, i.e. AB for the coarse case equals AB for the fine case. It is important that the first point be correctly chosen as the cyclic nature of the pattern results in an infinite number of intersections. In order to achieve this the values of AB for the coarse and fine cases are calculated using an initial arbitrarily small value of $S/2$ and incremented suitably until the curves cross. A suitable starting value is 1 metre. From this, using the cubic equation, the value of n is worked out for the upper and lower antennas, using the k and t values appropriate to the fine zone structure case. From these n values the proportion of the zone spanned for the fine structure follows (α_f) and the signal strength is determined using the above equation (2).

The values of k and t are now changed for the coarse case and the process repeated. With such low values of S the coarse case signal will be very low and the fine case approaching +6 dB. S is now incremented, say by 1 metre, and the process repeated: this continues with successive S increments until the coarse case signal exceeds the fine.

The cubic equation has to be solved for the upper and lower antennas for fine and coarse patterns, i.e. four times, for each S value. The prospect of attempting this by hand for more than a few S values is, to say the least, daunting. The iterative approach lends itself, however, to solution by a programmable calculator and the TI58 and TI59 have proved very well suited to this application. There are obvious advantages in using the magnetic card TI59, but care has been taken in compiling the software to ensure that it will be accommodated in the less capacious TI58. Although about 300 steps require to be entered manually into the TI58 this has not caused any problems and such a machine has been used on many occasions to run sets of these calculations. The printer, which can be used with either model, enables intermediate results to be listed, such as the coarse and fine zone minimum signal strengths at all spacing values. The programs given here, however, do not include printing instructions as these can readily be inserted by the user in accordance with his or her own requirements. The system runs perfectly satisfactorily without the printer.

6 Structure of the Program

The main constituent of the program is the algorithm for the CCITT cubic equation, in which the Fresnel zone numbers are calculated as required for the various stages of the process. This is assembled as a sub-routine (SBRA) and may be called many times during execution. A separate program for running this algorithm on its own, useful for general planning and analysis, is included as Appendix 1.

After the program has run the value of n is displayed, and the value of the received signal relative to f.s.r. can then be calculated by keying R/S. The display of signal strength is in decibels relative to f.s.r., negative numbers being below f.s.r., positive numbers being above (reinforcement) up to a maximum of approximately 6·02. As with the main program this algorithm is suitable for use with paths above grazing incidence.

In the main program, given in full as Appendix 2, sub-routine A is nested within a second, SBRB, which is called to calculate the values of n for the upper and lower antennas. During the execution of SBRB, SBRA is called twice. The flowchart for this sub-routine is shown as Fig. 7.

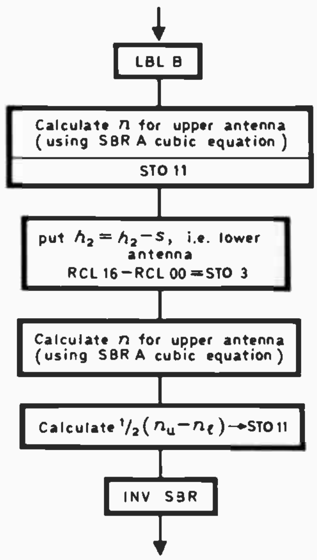


Fig. 7. Subroutine B flowchart.

In order that the method derived from Fig. 6 can be applied, values for the coarse and fine zone proportions spanned by the value of S being considered need to be calculated. This is accomplished by solving the cubic equation to establish the Fresnel zone number by calling SBRA, firstly for antenna height h_2 (giving n_u) and then for the lower antenna height of $(h_2 - S)$, which gives n_l. Following this $\frac{1}{2}(n_u - n_l)$ is calculated to become α in equations (1) and (2) above.

Values of S are incremented in 1 metre steps for successive iterations until the optimum spacing is exceeded. At this point S is decremented in 0·1 metre steps until the optimum point is passed once again. The program execution terminates here, the resultant S-value being within 0·1 metre of optimum. This is illustrated in detail by the flowchart given in Fig. 8.

Two pause instructions have been included, one in the 1 metre increment loop and one in the 0·1 metre decrement loop, such that the display flashes for half a second once every program cycle showing the separation being used. This occurs once every 35 to 40 seconds. Print instructions can conveniently be inserted here, and if they are also added to print out STO 11 at the appropriate times, the deepest fades due to the fine and the coarse patterns can be printed against the separation. These would be as a linear proportion of f.s.r., of course,

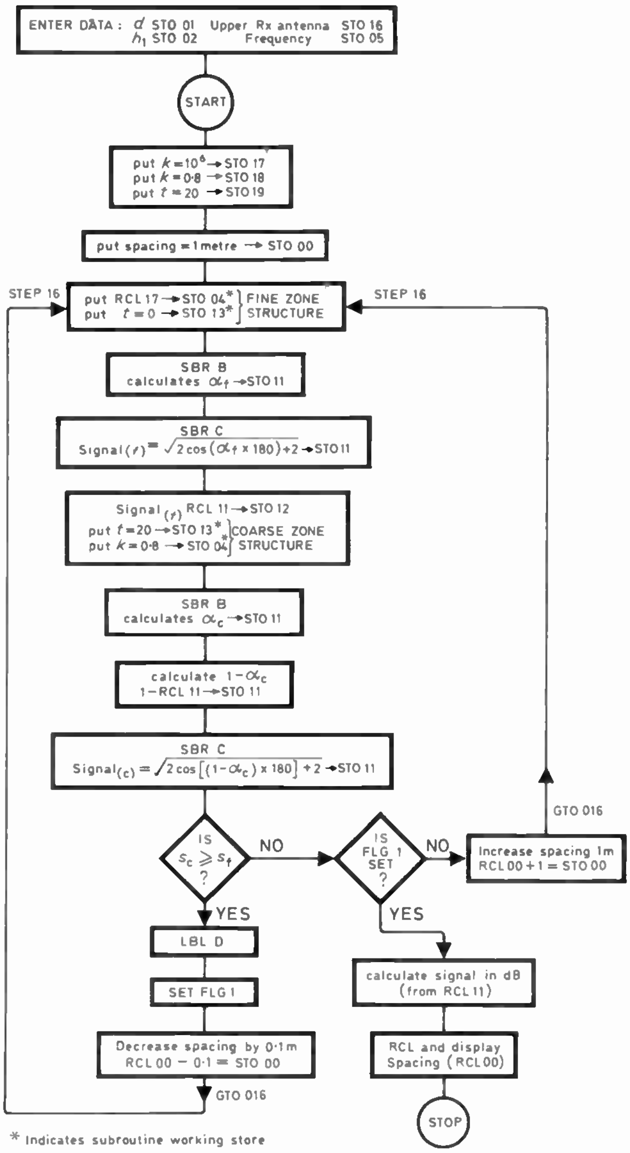


Fig. 8. Main diversity program flowchart.

and not dB as the dB calculation is only done at the end of the program. Finally, the signal strength during the deepest fade at the optimum *S* value is calculated, in decibels relative to f.s.r., and the result stored in STO 11.

It is not always possible to realize in practice the diversity spacing derived by the program. An example (based on a real North Sea 60/72 channel main bearer) is the case of a coast station on high ground serving an offshore platform much closer to sea-level. Although the offshore spacing was quite practical, the required onshore spacing exceeded the height of the mast at the coast station. In a case like this it is important to know the penalties of having to mount the antennas at non-optimum spacings. By simply running the main part of the program once with the practical spacing and choosing the lower signal strength (i.e. worst case) the deepest fade possible relative to f.s.r. is determined. To avoid the difficulty of having to break into and out of the program, a further routine has been written solely for this purpose into which the real-life values are inserted and which runs for about 8 minutes, stopping with the lowest possible signal displayed in dB relative to f.s.r. for any chosen *k*-value and over the range of *t* from 0–20 m. This is included as Appendix 3. The run can be repeated with a different *k*-value in STO 4, and so on until an adequate *k*-range has been explored. At the end of each run RCL 12 gives the value of *t* at which the deepest fade occurs. The flow chart of this program is given as Fig. 9.

7 Use of the Program

Loading programs into the TI58 and 59 is straightforward as owners of these machines will appreciate. The programs in the Appendices are written out ready for direct literal entry, following the switch-on instruction. It is always safest to start from machine switch-on, so that starting conditions are standardized.

Taking Appendix 2 as an example the first entry after switch-on is 2 2ND OP 17 (to partition the machine) then the LRN key to select the 'learn' condition, then the keystrokes exactly as written until the final LRN is keyed. The LRN operation is shown within a rectangular outline simply to indicate the extent of the 'learn' condition: each keystroke between these signs is written into the program stores. The small boldface numbers at the end of each line are not part of the program and must not be entered. They show the step number as displayed on the calculator at the end of each line as written, and act as a check on the accuracy of program entry to that point. Whereas the correct number does not guarantee accuracy, any discrepancy certainly indicates a mistake.

After the final LRN has been keyed the data can be entered, the check program forming a useful example. Then key CLR and RST to initialize the machine, followed by R/S to start the run. The machine does the rest. In the case of the check data given, the display at the end of the run (10 to 11 minutes) will show 12·7, which is the optimum diversity spacing in metres. Keying RCL 11 will give the minimum received signal level of −5·03 dB, indicating that the deepest fade possible is 5 dB below f.s.r. (it will sometimes be found that the minimum fade is actually above f.s.r.). Data for the next run may now be entered into store as before, then CLR–RST–R/S to run.

Perhaps a final note about the significance of these results is in order, particularly in a case where the deepest fade is above f.s.r., as in the test program in Appendix 3. Clearly a condition in which there is no reflection will result in no two-path fading and a constant f.s.r. signal. As the coefficient of reflection increases to −1 the signal in the Appendix 3 example will increase to +5·07 dB at the deepest fade and +6·02 dB at the maximum signal condition.

Where the deepest two-path fade is above f.s.r. the system needs to achieve a satisfactory noise characteristic

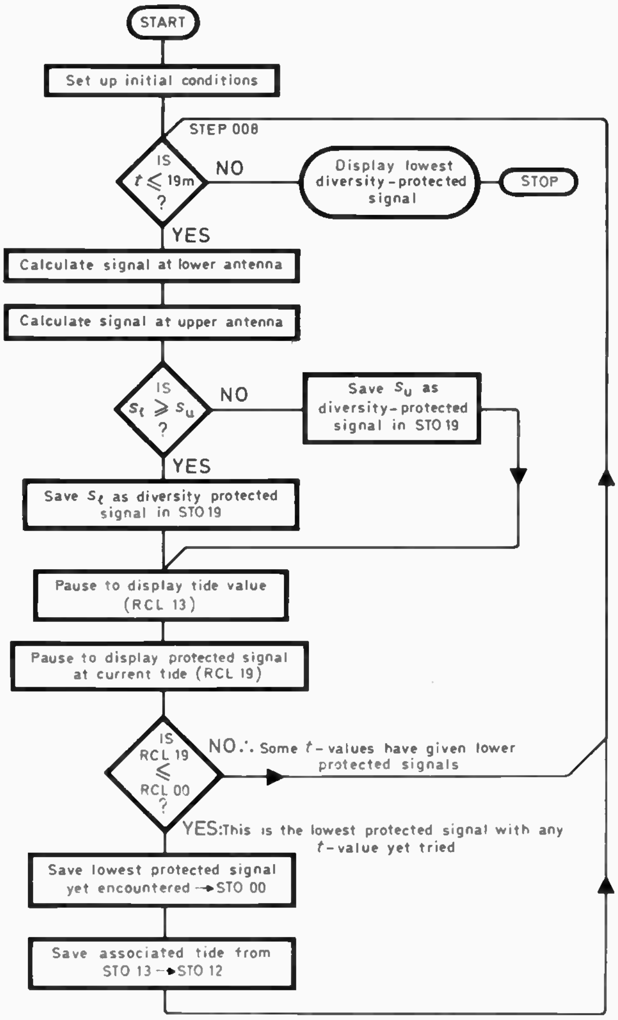


Fig. 9. Flow for Appendix 3 program.

at f.s.r.—this now being the weakest signal case. This is not usually difficult within normal line-of-sight planning parameters and with the S value calculated as described. When achieved, this represents all that is necessary to provide adequate protection against two-path fading.

If the data from the test program in Appendix 2 are run in the Appendix 3 program it will be found that the deepest fade is −4·03 dB ref. f.s.r., that is 1 dB stronger than the prediction of the Appendix 2 calculation. This discrepancy arises because the worst case phase relationship shown in Figs. 5 and 6 between the coarse and fine patterns does not necessarily occur in practice. A run through the Appendix 3 program after S has been determined will establish how much better the practical case is than the worst case assumption.

8 Conclusions

A program is appended which runs on a TI58 or TI59 pocket calculator and which derives the optimum diversity spacing for a line-of-sight radio link which:

Operates above grazing incidence
Uses dual space diversity
Uses either switched diversity or optimal ratio baseband combining
Operates over water, or otherwise such as to invoke the two-path propagation model.

In addition this program also gives a figure for the deepest fade possible at the derived spacing when the reflection coefficient is −1.

A second program enables the actual link performance to be predicted in terms of minimum received signal level when the reflection coefficient is −1 for use when the optimum spacing calculated from the first program cannot be realized due, for example, to physical restraints on site.

A third program enables the Fresnel zone number to be determined for a single antenna transmitter–receiver system together with the received signal strength corresponding to that zone number when the reflection coefficient is −1. It is valid for positive path clearances only.

Examples are given with each program to illustrate the running of the program and to check the programs for accurate entry prior to entering data relating to the problem(s) in hand.

These programs have been used in the planning of North Sea multichannel radio systems which have used frequencies in the L-band: however, the principle can be used at any frequency when two-path propagation obtains. The choice of the TI range of machines has been influenced by their relatively low cost and remarkable flexibility. They are machines widely used by radio engineers and the ability to store programs on magnetic cards together with the ready availability of a thermal printer makes for a most cost-effective and convenient way of solving these problems.

9 Acknowledgments

This paper is published with the permission of Brown & Root (U.K.) Ltd., and the author would like to record his grateful acknowledgment for the encouragement he has received from the Brown & Root Management. Encouragement has also come from the British Telecom International North Sea Task Force whose helpful comments have proved very valuable during this work. Finally—but by no means least—the author owes a particular debt to his colleague Mike Callis whose patience during many hours of calculations resulted in a particularly important contribution during the conceptual stages of the work.

10 References

1 Special Issue on Communications with Off-Shore Installations, *The Radio and Electronic Engineer*, **50**, no. 8, August 1980.
2 Post Office Telecommunications Headquarters, Specification RC 4031, '12/24 Telephone Channel Radio Equipment Operating in the Frequency Bands 1450–1535 MHz and 1700–1900 MHz'.
3 Post Office Telecommunications Development Department, Specification RC 4030, for Overall Performance Requirements and Supervisory Requirements for 72/132 Telephone Channel Radio-Relay Equipment for use in an Offshore Communication Network Operating in the Band 1700–1900 MHz'.

4 Post Office North Sea Telecommunications Services, 'Post Office Code of Practice for Planning Offshore Radio Links Serving Oil and Gas Production Platforms. NSTF 26/NSC 38, Appendix A.
5 Vigants, A., 'Space-diversity performance as a function of antenna separation', *IEEE Trans. on Communications Technology*, **COM-16**, no. 6, pp. 831–36, December 1968.
6 Lewin, L., 'Diversity reception and automatic phase correction', *Proc. Instn Elect. Engrs*, **109**, B, no. 56, pp. 295–304, July 1962.
7 C.C.I.T.T. 'Economic and Technical Aspects of the Choice of Transmission System', Appendix: Propagation.
8 Vigants, A., 'Space-diversity engineering', *Bell Syst. Tech. J.*, **54**, no. 1, pp. 103–42, January 1975.
9 Baker, A. E. and Brice, P. J., 'Radio-wave propagation as a factor in the design of microwave relay links', *P.O. Elect. Engrs J.*, **62**, pp. 239–46, January 1970.
10 Laine, R. U., 'Microwave propagation and reliability', *Automatic Electric Tech. J.*, **9**, part 4, pp. 142–50, October 1964.
11 Inoue, T. and Akiyama, T., 'Propagation characteristics on line-of-sight over-sea paths in Japan', *IEEE Trans. on Antennas and Propagation*, **AP-22**, no. 4, pp. 557–65, July 1974.
12 Bullington, K., 'Radio propagation fundamentals', *Bell Syst. Tech. J.*, **36**, no. 3, pp. 593–626, May 1957.
13 Millington, G., 'Curved Earth geometrical optics', *Marconi Rev.*, **80**, pp. 1–12, January–March 1946.
14 Sharman, J. R., 'Pre-detection combining', *Point-to-point Commun.*, **17**, no. 3, pp. 120–28, September 1973.
15 Pearson, K. W., 'Method for the prediction of the fading performance of a multi-section microwave link,' *Proc. IEE*, **112**, no. 7, pp. 1291–1300, July 1965.

11 Appendix 1: Fresnel Zone and Signal Strength for Given Antenna Heights

Displayed Step No.

Switch on, [LRN]
RCL 2 − RCL 13 = STO 14 RCL 3 − RCL 13 = STO 15 16
+ RCL 14 = × 8·5 × RCL 4 ÷ 4 + RCL 1 x^2 ÷ 12 = 37
STO 6 RCL 15 − RCL 14 = × RCL 4 ÷ 4 × RCL 1 × 6·37 = 59
÷ RCL 6 y^x 1·5 = INV 2nd COS = ÷ 3 + 240 = 2nd 77
COS × RCL 6 $\sqrt{x}$ × 2 + RCL 1 ÷ 2 = STO 7 +/_ + RCL 1 95
= x^2 × 4 ÷ 51 ÷ RCL 4 +/_ + RCL 15 = × (RCL 14 115
− RCL 7 x^2 × 4 ÷ 51 ÷ RCL 4) × RCL 5 ÷ 75 ÷ RCL 1 136
= STO 00 R/S +/_ + 1 = × 180 = 2nd COS × 2 + 2 155
= $\sqrt{x}$ 2nd LOG × 20 = STO 11 R/S 165
[LRN]

ENTER DATA → CLR → RST → R/S
Display: after execution contains Fresnel zone number.
Then key R/S:
Display: contains signal strength in dB ref. f.s.r.

Data		Check Values
d (km)	→ STO 01	5
h_1 (m)	→ STO 02	50
h_2 (m)	→ STO 03	28
k	→ STO 04	0·8
f (GHz)	→ STO 05	1·5
t (m)	→ STO 13	20

Check answer: n = 0·898758608
Signal = 5·910298925 dB (+ve).

12 Appendix 2: Optimum Diversity Spacing for Dual-switched Diversity Sea Path l.o.s. Links

K range 0·8 to 10^6; t range 0–20 m

Displayed Step No.

Switch on,
2 2nd OP 17 [LRN] 1 EE 6 STO 17·8 STO 18 20 11
STO 19 1 STO 00 RCL 17 STO 04 0 STO 13 BC STO 12
RCL 19 29
STO 13 RCL 18 STO 04 B 1 − RCL 11 = STO 11 C
RCL 12 46
$x \rightleftarrows t$ RCL 11 2nd $x \geq t$ D 2nd Iffig 1 E RCL 00 56
+ 1 = STO 00 2nd PAUSE GTO 16 2nd LBL D 2nd 67
Stflg 1 RCL 00 − ·1 = STO 00 2nd PAUSE GTO 16 79
2nd LBL B (RCL 16 STO 3 A STO 11 RCL 16 − RCL 00 96
= STO 3 A RCL 11 − RCL 14 = ÷ 2 = STO 11) INV SBR 113
2nd LBL C RCL 11 × 180 = 2nd COS × 2 + 2 = $\sqrt{x}$ 129
STO 11 INV SBR 2nd LBL A (RCL 2 − RCL 13 = STO 142
14 RCL 3 − RCL 13 = STO 15 + RCL 14 = × 8·5 × 160
RCL 4 ÷ 4 + RCL 1 x^2 ÷ 12 = STO 6 RCL 15 − 177
RCL 14 = × RCL 4 ÷ 4 × RCL 1 × 6·37 = ÷ RCL 196
6 y^x 1·5 = INV 2nd COS = ÷ 3 + 240 = 2nd COS 213
× RCL 6 $\sqrt{x}$ × 2 + RCL 1 ÷ 2 = STO 7 +/_ + RCL 1 230
= x^2 × 4 ÷ 51 ÷ RCL 4 +/_ + RCL 15 = × (RCL 14 250
− RCL 7 x^2 × 4 ÷ 51 ÷ RCL 4) × RCL 5 ÷ 75 ÷ RCL 271
1 = STO 14) INV SBR 2nd LBL E RCL 11 2nd LOG 282
× 20 = STO 11 RCL 00 R/S 291
[LRN]

ENTER DATA → CLR → RST → R/S
Display: During execution flashes S value being tried once per cycle, approx. every 35 seconds.
After execution contains S in metres.
RCL 11 to give max. fade in dB ref. f.s.r.

Data		Check Values
d (km)	→ STO 01	56·14
h_1 (m)	→ STO 02	301
h_{2b} (m)	→ STO 16	48·5
f (GHz)	→ STO 05	1·8

Check answers:
S = 12·7 m (displayed as $1{\cdot}27 \times 10^1$)
Max. fade = −5·03 dB ref. f.s.r.
Check program runs approx. 10 minutes.

13 Appendix 3: 1. Lowest dual-diversity signal over 0–20 m tide. 2. Tidal level at which 1 occurs

Displayed Step No.

Switch on, 2 2nd OP 17, [LRN]
10 STO 00 1 +/_ STO 13 RCL 13 $x \rightleftarrows t$ 19 2nd $x \geq t$ E
RCL 00 17
R/S 2nd LBL E RCL 13 + 1 = STO 13 C RCL 13
2nd PAUSE RCL 19 33
2nd PAUSE $x \rightleftarrows t$ RCL 00 2nd $x \geq t$ D GTO 008 2nd
LBL D 44
$x \rightleftarrows t$ STO 00 RCL 13 STO 12 GTO 008 2nd LBL C
(RCL 16 59
= STO 3 A STO 18 RCL 17 STO 3 A $x \rightleftarrows t$ RCL 18 2nd 73
$x \geq t$ B RCL 11 STO 19) INV SBR 2nd LBL B STO 19) 86
INV SBR 2nd LBL A (RCL 2 − RCL 13 = STO 14 RCL 3 −
RCL 13 103
= STO 15 RCL 1 x^2 ÷ 12 + 8·5 × RCL 4 ÷ 4 × (RCL 14 125
+ RCL 15) = STO 6 RCL 4 ÷ 4 × RCL 1 × 6·37 × (RCL 147
15 − RCL 14) = STO 7 (RCL 7 ÷ RCL 6 y^x 1·5) 166
INV 2nd COS = STO 8 RCL 6 $\sqrt{x}$ × 2 × (RCL 8 179
÷ 3 + 240) 2nd COS + RCL 1 ÷ 2 = STO 9 RCL 1 197
− RCL 9 = STO 10 ((((RCL 14 − RCL 9 x^2 × 4 216
÷ 51 ÷ RCL 4) × (RCL 15 − RCL 10 x^2 × 4 ÷ 51 ÷ RCL 238
4) × RCL 5 ÷ 75 ÷ RCL 1 +/_ + 1) × 180) 2nd COS ×
2 + 2) $\sqrt{x}$ 265
2nd LOG × 20 = STO 11) INV SBR 274
[LRN]

ENTER DATA → CLR → RST → R/S.

Data		Check Values
d (km)	→ STO 01	56·1
h_1 (m)	→ STO 02	328
h_{2a} (m)	→ STO 16	34·5
h_{2b} (m)	→ STO 17	48·5
k	→ STO 04	0·8
f (GHz)	→ STO 05	1·8

Running time: 8 minutes approx.
Check answer: +5·07 dB at 11 m tide.
Display: During execution: Flashes tide, then dB relative to free space at that tide (every 20–25 s).
At conclusion: Displays lowest dual diversity signal received over 0–20 m tide in dB ref. f.s.r.
RCL 12 gives tide at lowest signal level.

Manuscript first received by the Institution 30th October 1980 and in revised form on 10th February 1981
(Paper No. 2115/Comm 326)

UDC 621.397.331
Indexing Terms: Television systems, Digital coding

Adaptive discrete cosine transformation of pictures using an energy distribution logarithmic model

W. C. WONG, B.Sc., Ph.D.*

and

R. STEELE, B.Sc., Ph.D., C.Eng., M.I.E.E.*

SUMMARY

Adaptive transform coding using the discrete cosine transform for monochrome pictures is considered. Selecting which coefficients to transmit and with how many bits involves an estimate of each coefficient's energy. Dividing the image of 256 lines, 256 pels/line into blocks of 16×16 pels, we consider two estimation methods for the bit allocation process. One method involves the mean energy of each coefficient, while the other is an approximation to this method, called Energy Distribution Logarithmic Model, that drastically reduces the amount of side information that must be transmitted. To this we add an adaptive coefficient selection procedure based on the energy of three neighbouring coefficients. Performing the experiments by means of computer simulation we were able to achieve an average of 0·55 bits/pel for a recovered image having a mean square error of 0·1%.

* *Bell Laboratories, Crawford Hill Laboratory, Holmdel, NJ 07733.*

1 Introduction

Digital encoding of picture signals usually commences by sampling the raster scan image above the Nyquist rate to yield a two-dimensional array of samples, known as picture elements or pels. These pels can then be processed into binary signals by a variety of methods[1] which fall into two main classes: those which achieve bit-rate reduction in the time domain and those which gain it in a transformed domain. Currently the most popular exponent of the first category is adaptive differential pulse code modulation (a.d.p.c.m.) where the adaptation occurs in both the prediction and quantization processes and is subject to perceptual criteria.[2,3] With adaptive transform coding (a.t.c.), the generic name of coders in the second category, the pels are transformed, a block at a time, and bit-rate reduction is achieved by exploiting the properties of the transformed coefficients.[4] A two-dimensional a.t.c. system employing energy estimation criteria and an adaptive coefficient selection strategy is to be described and its performance evaluated by means of computer simulation. However, before doing so we will briefly state the essentials of a.t.c. as it effects our system.

2 Adaptive Discrete Cosine Transform Coding

There are numerous transforms that can be used for a.t.c. of images. However, to have any practical significance they should have three important characteristics: orthogonality, an efficient computation algorithm, and for a given bit-rate (achieved by discarding some coefficients and quantizing the remainder with differing accuracy), an acceptable perceptual quality. Many transforms have been exhaustively studied,[4] but the one that seems to have emerged with the best all-round performance is the discrete cosine transform (d.c.t.).[5] Interestingly it is also the one preferred in speech coding.

Consider a square segment of the image composed of M lines, containing M pels per line. Let the horizontal and vertical co-ordinates in the spatial image and the transformed image be x, y and m, n, respectively. The two-dimensional d.c.t. of the pel array

$$f(x, y);\ x, y = 0, 1, \ldots, M-1,$$

is the coefficient array $a(m, n);\ m, n = 0, 1, \ldots, M-1$, given by

$$a(m,n) = \frac{2}{M} C(m)C(n) \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} f(x,y) \times \times \cos\left[\frac{m\pi}{M}\left(x+\frac{1}{2}\right)\right] \cos\left[\frac{n\pi}{M}\left(y+\frac{1}{2}\right)\right] \quad (1)$$

and its inverse, i.e. the spatial image, is

$$f(x,y) = \frac{2}{M} \sum_{m=0}^{M-1} \sum_{n=0}^{M-1} C(m)C(n)a(m,n) \times \times \cos\left[\frac{m\pi}{M}\left(x+\frac{1}{2}\right)\right] \cos\left[\frac{n\pi}{M}\left(y+\frac{1}{2}\right)\right] \quad (2)$$

0033–7722/81/110571+08 $1.50/0

where

$$C(0) = \frac{1}{\sqrt{2}} \text{ and } C(1), C(2), \ldots, C(M-1) = 1.$$

As most pictures contain significant correlation over a distance of about 20 pels, a good choice for M is 16. Too small a value of $M(<4)$ prevents a.t.c. from exploiting the correlative properties of the image.

The a.t.c. system must adaptively select and quantize the coefficients $a(m, n)$; $m, n = 0, 1, \ldots, 1-M$. Although the variation in the amplitudes of the coefficients in the block is extremely wide compared to the variation in the amplitudes of the pels in the block, only a few coefficients have large values. As the total image energy in the spatial and transform domains are identical, namely,

$$E_T = \sum_{m=0}^{M-1} \sum_{n-0}^{M-1} a(m, n)^2 = \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} f(x, y)^2 \qquad (3)$$

it follows that many coefficients have negligible values and can often be set to zero without causing serious degradation in the inversed transformed, i.e. recovered, image. It is the ability to set some coefficients to zero, and thereby avoid their transmission, that significantly contributes to bit-rate reduction in a.t.c. Thus the M^2 pels are transformed to M^2 coefficients, those with insignificant values rejected, and the remainder transmitted in a binary format. The receiver pads the missing coefficients with zeros (although estimates of their values can be made) to form a M-by-M array which is inverse transformed to recover a M-by-M block of pels that is required to be a good approximation to the original block.

The variation of the logarithm to base ten of the energy of the coefficients along the m and n co-ordinates for the 'Girl' picture of Fig. 1 is shown in Fig. 2. Observe that the energy in the transform domain is concentrated in a relatively small number of coefficients associated with $m = n = 0$, and other low values of m and n. For high values of m and n, particularly those close to $M-1$, the coefficients are small and can generally be discarded. Thus, although the correlation of the amplitudes of the coefficients is relatively low (compared to that for the pels in the spatial image), the correlation of their energies is substantially higher. This feature is exploited in the coefficient selection and bit assignment procedures.

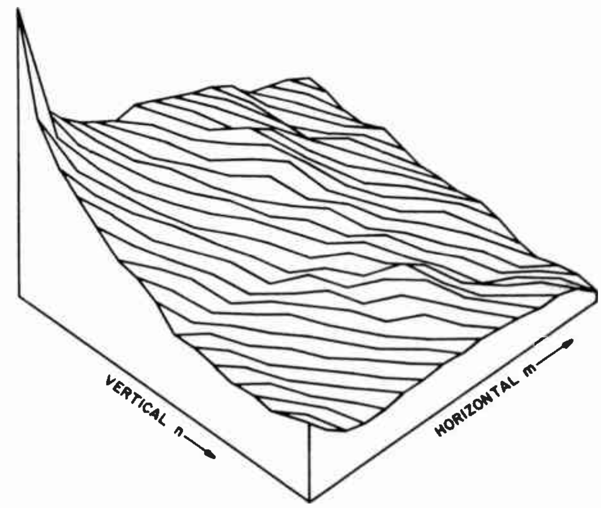


Fig. 2. The logarithm of the mean energy ($\tilde{E}(m, n)$) distribution of the two-dimensional d.c.t. coefficients for the 'Girl' picture of Fig. 1.

In a.t.c. we must decide which coefficients are necessary for the reconstructed picture quality, and how they are to be quantized. For quantization to occur we must determine the quantizer's characteristic, the scaling factor for the coefficients to be quantized and the number of quantization levels to be used, i.e. the number of bits to represent the quantized coefficients. Selection and quantization are not independent. The coefficient $a(m, n)$, whose energy (or its estimate) is $E(m, n)$, is quantized by a quantizer whose levels are represented by binary code words having $b(m, n)$ bits. Our first step is to consider how to determine $b(m, n)$.

Fig. 1. The 'Girl' picture.

2.1 Bit Assignment

From rate distortion theory[6] the bits allocated to the sample at the output of a memoryless Gaussian source at the jth instant is

$$b(j) \geqslant \frac{1}{2}\log_2\left(\frac{\sigma^2(j)}{D}\right) \text{ bits}; \quad \sigma^2(j) \geqslant D$$

and (4)

$$b(j) = 0; \quad \sigma^2(j) < D$$

where $\sigma^2(j)$ is the variance of the source and D the average distortion power.

In a.t.c. the bit assignment $b(m, n)$ of $a(m, n)$ is based on equation (4), and is

$$b(m, n) = \text{INT}\left[\frac{1}{2}\log_2\left(\frac{E(m, n)}{D_1}\right)\right] \qquad (5)$$

where

$$D_1 = \frac{E_{mean}}{k_1} \quad (6)$$

$$E_{mean} = \frac{1}{XYM^2} \sum_{x=0}^{MX-1} \sum_{y=0}^{MY-1} f^2(x, y) \quad (7)$$

X and Y are the number of blocks in the image along the x and y co-ordinate, where each block has M^2 pels, and k_1 is a system parameter that determines the quality of the reconstructed image.

Comparing equations (4) and (5) we see that the term INT $[\cdot]$ meaning integer of $[\cdot]$ is introduced as we only consider quantizers having integer bits. The variance term $\sigma^2(j)$ is replaced by the energy expression $E(m, n)$ as D_1 is also a function of the mean energy of the picture.

The number of bits $b(m, n)$ depends on the ratio of the coefficient energy $E(m, n)$ to the average distortion D_1. By arranging for D_1 to be proportional to the mean energy E_{mean} of the entire image, the mean square error in the decoded image becomes a function of E_{mean}. This is an acceptable criterion as it allows a high mean square error to prevail in pictures having high energy and vice versa, thereby maintaining a nearly constant signal-to-distortion ratio for different images.

The choice of $E(m, n)$ is clearly crucial as it not only assigns the number of bits $b(m, n)$ to quantize coefficient $a(m, n)$, but it also determines if $a(m, n)$ is to be selected for transmission.

2.2 Quantization Process

The positive value of $b(m, n)$ determines the number $L(m, n)$ of quantization levels as

$$L(m, n) = 2^{b(m,n)}. \quad (8)$$

The quantization characteristic is designed according to a model based on the statistics of $a(m, n)$. Figure 3 shows the histograms of the first eight transform coefficients for a block size of 16×16 pels and the 'Girl' picture of Fig. 1. From these histograms a Rayleigh p.d.f.

$$P(a(0,0)) = \frac{a(0,0)}{\sigma^2_{(0,0)}} \exp\left[-a^2(0,0)/2\sigma^2_{(0,0)}\right] \quad (9)$$

was assigned to the d.c. coefficient $a(0, 0)$, and a Gaussian p.d.f.

$$P(a(m,n)) = \frac{1}{\sqrt{2\pi\sigma^2_{(m,n)}}} \exp\left[-a^2(m,n)/2\sigma^2_{(m,n)}\right] \quad (10)$$

to all the other coefficients $a(m, n)$; $m, n = 0, 1, 2, \ldots, M-1$, except $m = n = 0$. $\sigma_{(0,0)}$ and $\sigma_{(m,n)}$ are the standard deviations of the Rayleigh and Gaussian p.d.f.s, respectively.

The variance of the d.c. coefficient $a(0, 0)$ is

$$V(0,0) = \int_0^\infty a^2(0,0) \times \left\{\frac{a(0,0)}{\sigma^2_{(0,0)}} \exp\left[-a^2(0,0)/2\sigma^2_{(0,0)}\right]\right\} da(0,0)$$

$$= 2\sigma^2_{(0,0)} \quad (11)$$

or,

$$\sigma_{(0,0)} = \sqrt{\frac{V(0,0)}{2}} = \sqrt{\frac{E(0,0)}{2}}. \quad (12)$$

By similar arguments the standard deviation of coefficient $a(m, n)$ is

$$\sigma_{(m,n)} = \sqrt{V(m,n)} = \sqrt{E(m,n)} \quad (13)$$

The quantization process operates as follows. The coefficient is scaled by its standard deviation and quantized by a Max quantizer[7] optimized for the appropriate distribution and having a number of levels

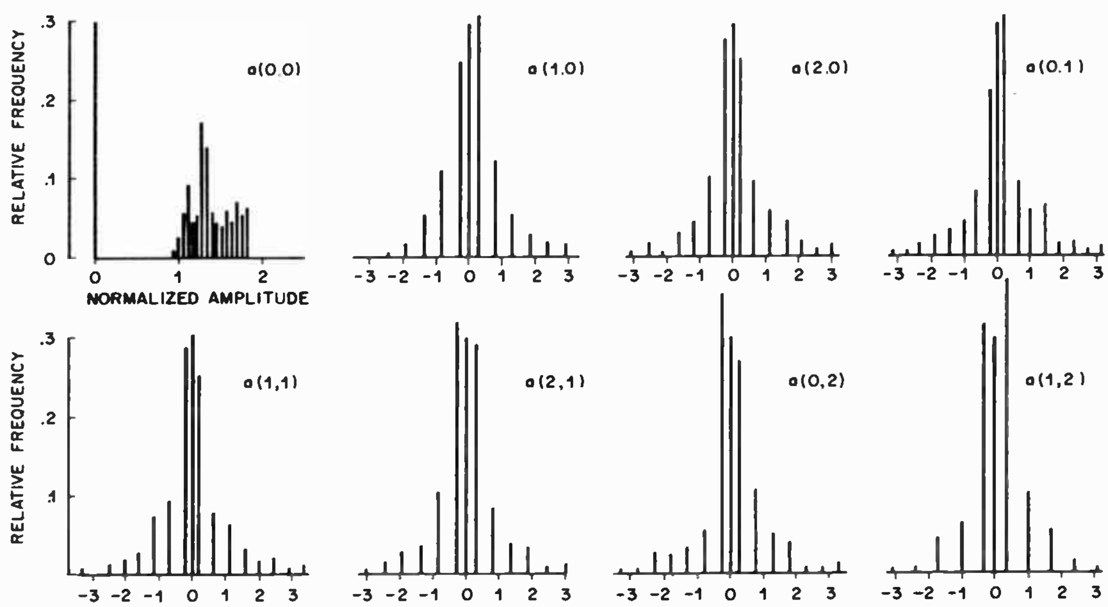


Fig. 3. Histograms of the first eight transform coefficients for the 'Girl' picture of Fig. 1. $M = 16$.

given by equation (8). For example, the coefficient $a(j,k)$ $(j,k \neq 0)$ is modified to $a(j,k)/\sigma_{(j,k)}$ prior to quantization by a quantizer whose characteristic is designed for $L(j,k)$ levels and for a Gaussian p.d.f. with a standard deviation of unity. The decision and reconstruction levels for the Gaussian quantizer for various number of bits are listed in Table 1. In our experiments the number of bits was limited to 8 for $a(0,0)$ and 6 for $a(m,n)$; $m, n = 0, 1, 2, \ldots, M-1$, except $m = n = 0$. Observe that the process of normalizing the coefficient amplitudes enables the effective quantization of coefficients having the same p.d.f. but widely differing energies.

The importance of the energy of the coefficient is apparent. It determines the bits assigned to the quantization process and the scaling of each coefficient,

Table 1

Placement of decision and reconstruction levels for Max Gaussian Quantizer having 1, 2, 3, 4, 5, 6 and 7 bits.

BITS	d_k	r_k
1	.00000	.79800
2	.00000	.45280
	.98160	1.51000
3	.00000	.24510
	.50060	.75600
	1.05000	1.34400
	1.74800	2.15200
4	.00000	.12840
	.25920	.38910
	.52240	.65680
	.79960	.94240
	1.09900	1.25600
	1.43700	1.61800
	1.84400	2.09600
	2.40100	2.73300
5	.00000	.06590
	.13200	.19810
	.26480	.33140
	.39910	.46680
	.53590	.60300
	.67610	.74730
	.82100	.89470
	.97180	1.04900
	1.13000	1.21200
	1.29900	1.38700
	1.48200	1.57700
	1.68200	1.78800
	1.90800	2.02900
	2.17400	2.31900
	2.50500	2.69200
	2.97700	3.26300
6	.00000	.03224
	.06696	.10050
	.13400	.16760
	.20130	.23500
	.26880	.30280
	.33680	.37100
	.40530	.43980
	.47450	.50930
	.54440	.57970
	.61620	.65100
	.68690	.72340
	.76010	.79710
	.83450	.87230
	.91060	.94920
	.98830	1.02800
	1.06800	1.10900
	1.15000	1.19200
	1.23500	1.27900
	1.32300	1.36800
	1.41500	1.46200
	1.51000	1.56000
	1.61100	1.66300
	1.71700	1.77300
	1.83000	1.89000
	1.95200	2.01600
	2.08300	2.15400
	2.22900	2.30800
	2.37000	2.48200
	2.57900	2.68500
	2.80200	2.93400
	3.08600	3.2650
	3.48900	3.79200
7	.00000	.01678
	.03368	.05055
	.06545	.08426
	.10110	.11800
	.13490	.15180
	.16880	.18570
	.20260	.21960
	.23660	.25360
	.27070	.28770
	.30480	.32190
	.33910	.35630
	.37350	.39080
	.40810	.42540
	.44280	.46020
	.47770	.49520
	.51280	.53040
	.54810	.58590
	.58370	.60150
	.61950	.63750
	.65550	.67360
	.69190	.71010
	.72850	.74690
	.76550	.78410
	.80280	.82160
	.84050	.85950
	.87960	.89780
	.91720	.93660
	.95620	.97590
	.99560	1.01600
	1.03600	1.05600
	1.07600	1.09700
	1.11700	1.13800
	1.15900	1.18000
	1.20200	1.22300
	1.24500	1.26700
	1.28900	1.31100
	1.33400	1.35700
	1.38000	1.40300
	1.42600	1.45000
	1.47400	1.49900
	1.52300	1.54600
	1.57400	1.59900
	1.62500	1.65200
	1.67800	1.70500
	1.73300	1.76100
	1.78900	1.81800
	1.84800	1.87800
	1.90900	1.94000
	1.97200	2.00400
	2.03800	2.07200
	2.10600	2.14200
	2.17900	2.21700
	2.25500	2.29500
	2.33600	2.37900
	2.42300	2.46980
	2.51600	2.56500
	2.61700	2.67100
	2.72800	2.78600
	2.85100	2.91600
	2.99100	3.06800
	3.15300	3.24600
	3.34900	3.46500
	3.60000	3.76100
	3.96300	4.23900

and is involved in selecting which coefficients are to be quantized. However, we will see that different estimates of $E(m,n)$ can be used in coefficient selection and bit assignment.

3 Mean Energy Estimation

The mean energy of each coefficient $a(m,n)$ in the block of M by M pels is

$$\tilde{E}(m,n) = \frac{1}{XY}\sum_{i=1}^{XY} a_i^2(m,n) \tag{14}$$

where XY is the number of blocks in the picture. Observe that $\tilde{E}(m,n)$ for $m,n = 0, 1, \ldots, M-1$ must be transmitted as side-information, and that $b(m,n)$ is the same for each block in the image. In Section 5 we will show that although $b(m,n)$ is the same for each block, we can use different coefficient selection schemes in each block so that the number of coefficients transmitted will differ from block to block.

4 Energy Distribution Logarithmic Model

This model is an attempt to approximate the mean energy estimate (m.e.e.) with a large reduction in transmitted side information. In forming the model we take cognizance that the logarithm of the energy over the m–n plane forms a surface that decreases from the $m = n = 0$ point, as m and n increases in a manner that is

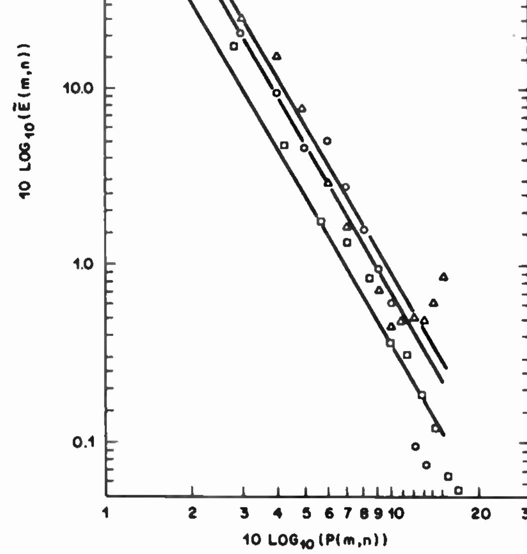


Fig. 4. The logarithm of the mean energy of the coefficients as a function of the logarithm of the distance factor P for the first row (○) and column (△), and the diagonal (□) of the coefficient plane.

approximately exponential. Inspecting the m–n plane, we selected the three paths, namely:

the first row $n = 0$, $m = 0, 1, \ldots, M-1$
the first column $m = 0$, $n = 0, 1, \ldots, M-1$
the diagonal $m = n = 0, 1, 2, \ldots, M-1$.

The variation of $\log_{10} \tilde{E}(m, n)$, where $\tilde{E}(m, n)$ is given by equation (14), was plotted as a function of

$$\log_{10} P(m, n) = \log_{10} \sqrt{m^2 + n^2}$$

for the coefficients in the first row and column and along the diagonal. The results are shown in Fig. 4. The points were found to correspond closely to a straight line, except at high values of $P(m, n)$, a region where the coefficient energy is small and consequently a region where coefficients are often discarded. We may consider that the row, column and diagonal paths are three straight lines on the surface of $\log_{10} \tilde{E}(m, n)$ as a function of $\log_{10} P(m, n)$. However, Fig. 4 shows that $\log_{10} \tilde{E}(m, n)$ differs in magnitude along these paths. We therefore propose a simple model to represent the asymmetrical nature of the surface, and to provide estimates of energy on any part of the surface, i.e. for any m and n. This model is represented by a distance measure $Q(m, n)$ given by

$$Q(m, n) = \sqrt{A_1(A_2 m^2 + A_3 n^2)} \qquad (15)$$

where A_1 is a normalizing factor given by

$$A_1 = \frac{1}{2}[\ln \tilde{E}(1, 0) + \ln \tilde{E}(0, 1)]$$

and A_2 and A_3 are the directional weighting factors

$$A_2 = 1/\ln(\tilde{E}(1, 0)) \qquad (16)$$

and

$$A_3 = 1/\ln(\tilde{E}(0, 1)). \qquad (17)$$

$\tilde{E}(1, 0)$ and $\tilde{E}(0, 1)$ are therefore employed to give an indication of the activity or energy distribution along the m and n axes. If $\tilde{E}(1, 0) = \tilde{E}(0, 1)$, $Q(m, n) = P(m, n)$.

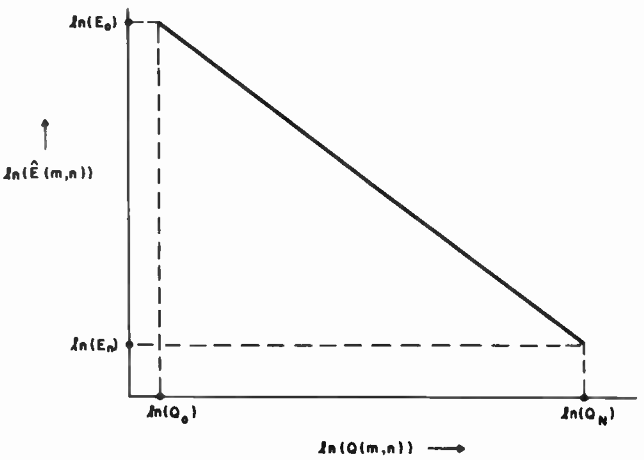


Fig. 5. Energy distribution logarithmic model (e.d.l.m.) equation.

The logarithm of the estimated energy, namely $\ln(\hat{E}(m, n))$, as a function of $\ln(Q(m, n))$ that represents the surface in the m–n plane, is given by the straight line equation

$$\ln(\hat{E}(m, n)) = C_1 \ln(Q(m, n)) + C_2 \qquad (18)$$

for $m, n = 0, 1, 2, \ldots, M-1$, except $m = n = 0$. C_1 and C_2 represent the slope and intercept of this linear log–log relationship whose end-points are specified by E_0, E_N, Q_0 and Q_N, as shown in Fig. 5. From the Figure,

$$C_1 = \frac{\ln(E_N) - \ln(E_0)}{\ln(Q_N) - \ln(Q_0)} \qquad (19)$$

and

$$C_2 = \ln(E_N) - C_1 \ln(Q_N). \qquad (20)$$

The values of E_0, E_N, Q_0 and Q_N must be selected to minimize the transmitted bit-rate for a given fidelity criterion. The criterion adopted here is a normalized mean square error (n.m.s.e.) of 0·1% which results in decoded pictures of good quality under 'normal' viewing conditions. The n.m.s.e. is defined by

$$\text{n.m.s.e.} \triangleq \frac{\sum_{x=0}^{XM-1} \sum_{y=0}^{YM-1} \{f(x, y) - F(x, y)\}^2}{\sum_{x=0}^{XM-1} \sum_{y=0}^{YM-1} f^2(x, y)} \qquad (21)$$

where $F(x, y)$ is the recovered image, and XM and YM are the number of pels in the x and y co-ordinates of the entire image, respectively.

The end-point E_0 is given by

$$E_0 = \max[\tilde{E}(1, 0), \tilde{E}(0, 1)] \qquad (22)$$

and Q_0 by

$$Q_0 = \begin{cases} \sqrt{A_1 A_3}; & E(0, 1) \geq E(1, 0) \\ \sqrt{A_1 A_2}; & E(0, 1) < E(1, 0) \end{cases} \qquad (23)$$

The maximum value Q_N is

$$Q_N = Q(M-1, M-1) \qquad (24)$$

and its corresponding E_N is a system parameter selected to minimize the bit-rate.

In using the energy distribution logarithmic model (e.d.l.m.) we select a value of E_N and in doing so find $\hat{E}(m, n)$ for any m and n. $\hat{E}(m, n)$ then replaces $E(m, n)$ in equation (5) and for a given k_1, $b(m, n)$ is calculated and quantization occurs as described in Section 2.2. To find E_N and k_1 we proceed as follows. E_N is kept constant and k_1 varied to yield a graph of n.m.s.e. as a function of bits/pel. The procedure is repeated for different values of E_N, and the E_N and k_1 finally selected are those which give the smallest n.m.s.e. for a given value of bits/pel.

For the 8-bit linear p.c.m. picture shown in Fig. 1, the a.t.c. using d.c.t., $M = 16$, $X = Y = 16$, the e.d.l.m. of equation (18), maximum values of $b(0, 0)$ and $b(m, n)$ of 8 and 6 bits, respectively, the best value of E_N was $\simeq 0{\cdot}08$.

The curve of n.m.s.e. versus bits/pel is shown in Fig. 6. From informal viewing experiences we concluded that when the normalized mean square error is 0·1% the degradation of picture quality is perceptible but not annoying. We therefore used this number as our objective measure of picture quality; and refer the reader to Fig. 9 for pictures having n.m.s.e. $\simeq$ 0·1%. From Fig. 6, the 'Girl' image can be reconstructed according to our quality criterion using 1·08 bits/pel, ignoring the transmission of side information.

When the mean energy estimate (m.e.e.) is used instead of e.d.l.m., the performance of the system improves as shown in Fig. 6, where for n.m.s.e. = 0·1% only 0·89 bits/pel are generated. This is not surprising as

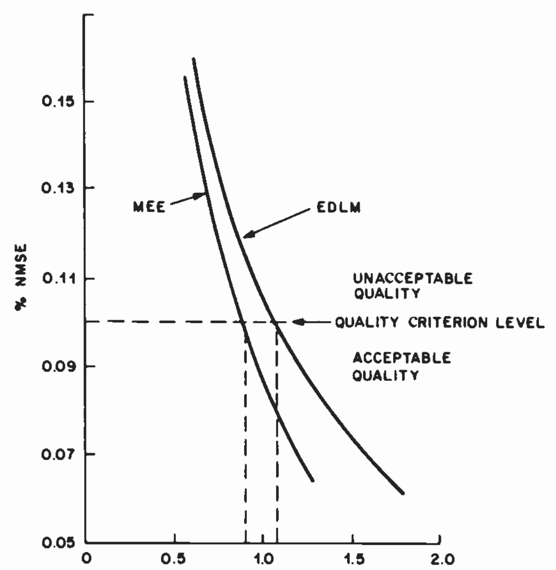


Fig. 6. Normalized mean square error (n.m.s.e.) as a function of bits per pel when the energy estimates are derived from the mean energy estimate procedure and the energy distribution logarithmic model.

the difference between the curves in Fig. 6 provides a measure of the accuracy of e.d.l.m., as e.d.l.m. is essentially an approximation to m.e.e. However, m.e.e. requires $M^2(=256)$ 8-bit code words representing $\tilde{E}(m, n)$ to be transmitted as side-information increasing the bits/pel to 0·92. To protect against transmission errors channel coding must be deployed. For example, if each 8-bit word of side information is triplicated the bits/pel to reconstruct the image is 0·98. The e.d.l.m. method requires only $\tilde{E}(1, 0)$ and $\tilde{E}(0, 1)$ (we assume E_N and k_1 are fixed) to be transmitted, a negligible increase in the bits/pel. Thus we observe that for n.m.s.e. = 0·1%, m.e.e. requires approximately 0·1 bit/pel less than e.d.l.m.

5 Adaptive Coefficient Selection

In the previous Sections, the m.e.e. and e.d.l.m. schemes were shown to provide an estimate of the energy of each coefficient to enable a bit allocation map to be formed. The map is the same for all blocks in the image. However, the activity in each block may be substantially different, and this offers the opportunity to achieve a reduction in the bit-rate by discarding coefficients having insignificant amplitudes, despite the fact that bits may have been allocated to them.

A simple adaptive selection strategy similar to that proposed by Wong and Steele[8] will now be described. It operates by exploiting the correlation in the distribution of transform coefficient energies, and initially the coefficients $a(0,0)$, $a(1,0)$, $a(0,1)$ and $a(1,1)$ are transmitted to start the adaptive coefficient selection procedure. An estimate $E'(m,n)$ is made of the energy of $a(m,n)$ and this is compared with a predetermined threshold related to the required fidelity of the decoded image. If

$$E'(m,n) \leqslant \frac{E_{\text{mean}}}{k_2} \tag{25}$$

where k_2 is a distortion threshold factor, the coefficient $a(m,n)$ is discarded. Further, all other higher-order coefficients along the row are also neglected. The procedure starts anew on the next row and so on until the block has been processed.

$E'(m,n)$ is formed as the average energy from three local coefficients as follows. For the first column,

$$E'(0,2) = \frac{1}{3}[\hat{a}^2(0,1)+\hat{a}^2(1,0)+\hat{a}^2(1,1)]$$

$$E'(0,3) = \frac{1}{3}[\hat{a}^2(0,1)+\hat{a}^2(0,2)+\hat{a}^2(1,1)] \tag{26}$$

$$E'(0,n) = \frac{1}{3}\sum_{j=1}^{3}\hat{a}^2(0,n-j); \quad n = 4, 5, \ldots, M-1.$$

and for the first row,

$$E'(2,0) = \frac{1}{3}[\hat{a}^2(1,0)+\hat{a}^2(0,1)+\hat{a}^2(1,1)]$$

$$E'(3,0) = \frac{1}{3}[\hat{a}^2(1,0)+\hat{a}^2(2,0)+\hat{a}^2(1,1)] \tag{27}$$

$$E'(m,0) = \frac{1}{3}\sum_{j=1}^{3}\hat{a}^2(m-j,0); \quad m = 4, 5, \ldots, M-1.$$

where the hat (ˆ) above the symbol signifies a quantized value. The first coefficient discarded along the first column or the first row, precipitates the abandonment of all subsequent coefficients along that column and row. For the remaining coefficients in the block $E'(m,n)$ is computed as

$$E'(m,n) = \frac{1}{3}[\hat{a}^2(m,n-1)+\hat{a}^2(m-1,n)+\hat{a}^2(m-1,n-1)] \tag{28}$$

Starting with $m = 1, n = 2, \ldots, M-1$, then $m = 2, n = 1, \ldots, M-1$, and so on until $m = M-1, n = 1, \ldots, M-1$. Whenever a coefficient is

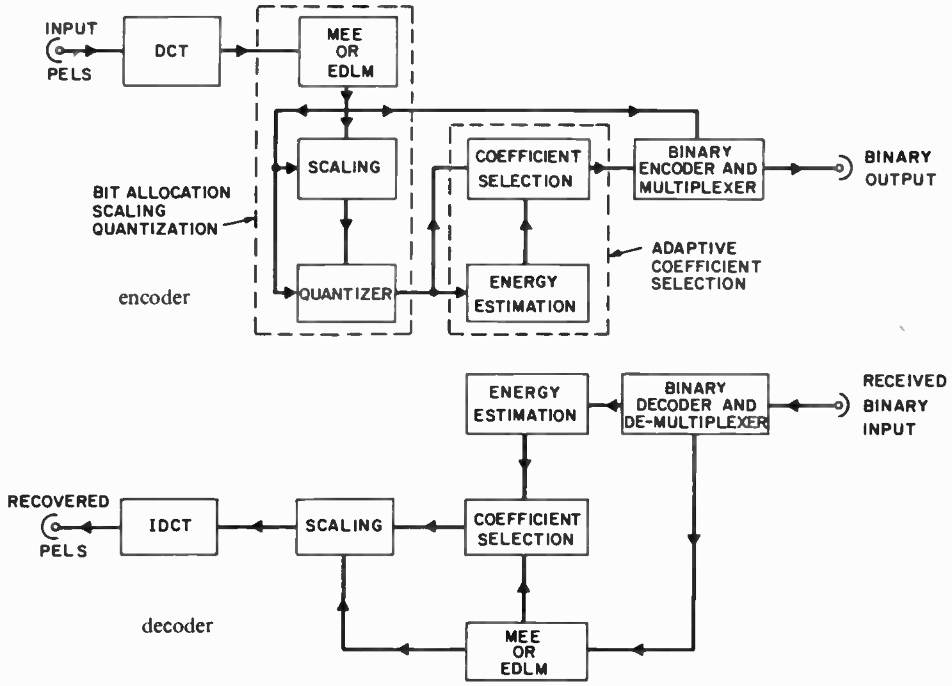


Fig. 7. Block diagram of m.e.e. or e.d.l.m. bit allocation scheme with a.c.s.

discarded all subsequent coefficients along the same row are also discarded.

6 M.E.E. and E.D.L.M. with Adaptive Coefficient Selection

In these schemes, designated m.e.e.-a.c.s. and e.d.l.m.-a.c.s., bit allocation (which includes the first coefficient selection process) and quantization occur as described in Sections 2, 3 and 4. This is followed by adaptive coefficient selection (a.c.s.) of the quantized coefficients according to equations (25) to (28). Assuming errorless transmission, identical energy estimates $E'(\cdot)$ to those used in a.c.s. at the encoder can also be employed at the receiver. The block diagram of the encoder and decoder are shown in Fig. 7.

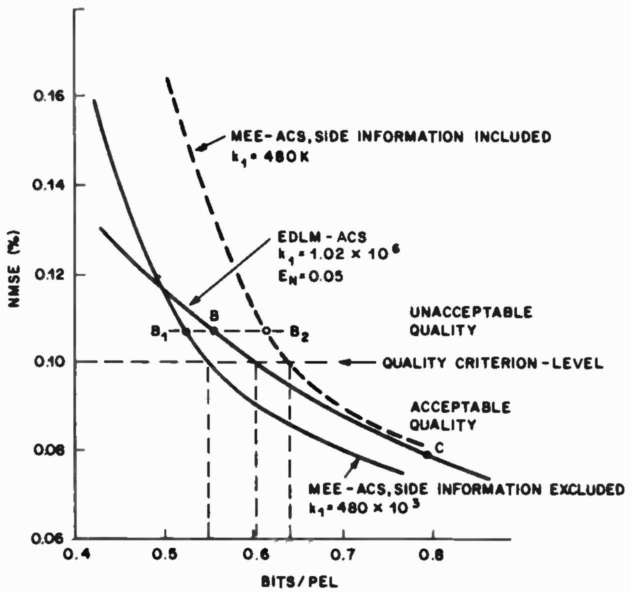


Fig. 8. N.m.s.e. as a function of bits/pel for m.e.e.-a.c.s. and e.d.l.m.-a.c.s. schemes.

Variation of n.m.s.e. as a function of bits/pel for both schemes is presented in Fig. 8 for the 'Girl' picture. k_1 was selected to give best performance, and the curves were obtained by varying k_2. M.e.e.-a.c.s. appears to require fewer bits per pel than e.d.l.m.-a.c.s. for n.m.s.e. of 0·1%. However, when the protected side information is included, the m.e.e.-a.c.s. has a curve similar to that shown dotted while that for e.d.l.m.-a.c.s. is substantially unchanged. Comparing Figs. 6 and 8 we see that the introduction of a.c.s. results in a reduction of nearly 50% in bits/pel at n.m.s.e. of 0·1% for e.d.l.m.

The photograph of the recovered image of the 'Girl' picture in Fig. 1, encoded with 8 bits/pel, is shown in Fig. 9 for the e.d.l.m.-a.c.s. scheme having $k_1 = 1{\cdot}02 \times 10^6$ and $k_2 = 32 \times 10^3$. We do not present photographs for the m.e.e.-.a.c.s. scheme as for a given n.m.s.e. they appear to be identical to those obtained by the e.d.l.m.-a.c.s. system. Thus the photograph of Fig. 9 also applies to m.e.e.-a.c.s. at points B_1 and B_2 in Fig. 8, depending on whether side-information is considered or not.

7 Conclusions

Adaptive transform coding using the discrete cosine transform (a.t.c.-d.c.t.) has been investigated by means of computer simulation, where different methods were employed to select and quantize the transformed coefficients for the image source shown in Fig. 1. A bit

Fig. 9. Recovered picture of the 'Girl' image of Fig. 1 obtained with e.d.l.m.-a.c.s. having $k_1 = 1{\cdot}02 \times 10^6$, $k_2 = 32 \times 10^3$, and a n.m.s.e. = 0·107%.

assignment method derived from rate distortion theory (eqn. (5)) requiring the estimation of the energy of each coefficient has been used. Exploiting the correlation in the energy of the two-dimensional coefficient array enabled us to produce two methods of energy estimation called mean energy estimation (m.e.e.), and its approximation version, the energy distribution logarithmic model (e.d.l.m.). The e.d.l.m. has the advantage compared to m.e.e. of requiring the transmission of negligible amount of side-information, although its performance in terms of bits/pel for a given n.m.s.e. is marginally inferior.

Using the m.e.e. and e.d.l.m. schemes, coefficient selection and bit allocation was the same for each block of M by M pels in the picture. To account for variations in the activity in different blocks we added the second process of adaptive coefficient selection (a.c.s.) which examined the energy of each coefficient relative to three of its neighbours. If the energy of the coefficient was below a threshold it was rejected together with all subsequent coefficients along its row. This had the result of a significant reduction in the number of transmitted coefficients for the same recovered picture quality. The amount of side information associated with the m.e.e. now became significant when a.c.s. was added and e.d.l.m.-a.c.s. was preferred achieving 0·55 bits/pel for n.m.s.e. of 0·1%.

The experimental results relate to the well-known 'Girl' image.[4] The statistical properties of the $M \times M$ blocks within the image are similar to those of other head and shoulder images, and consequently the results presented here may be viewed as typical of these types of pictures. Extrapolating to how the e.d.l.m.-a.c.s. system would perform in the context of video conferencing is contentious, but the perceived system noise may be mitigated by the effects of motion. The results given here, therefore, provide an approximate measure of what might be achieved if the e.d.l.m.-a.c.s. system is applied to moving head and shoulder images. Making performance comparisons with other encoding systems is difficult. A recent survey[9] of digital picture coding schemes pointed out that those with high complexity and low bits/pel have not been extensively studied with a large variety of pictures, and subjected to rigorous subjective testing. And so it is with our system. We can conclude that our scheme has similar complexity and performance with other recent techniques,[9] and that it does provide another interesting approach to low bit-rate transmission of images.

8 Acknowledgment

The authors thank J. E. Thompson and R. C. Nicol of the British Post Office for their support of this work, and D. J. Goodman for his constructive criticism during the preparation of the paper.

9 References

1 Netravali, A. N. and Limb, J. O. 'Picture coding: a review', *Proc. IEEE*, **68**, no. 3, pp. 366–406, March 1980.

2 Sharma, D. K. and Netravali, A. N. 'Design of quantizers for dpcm coding of picture signals', *IEEE Trans. on Communications*, **COM-25**, no. 11, pp. 1267–74, November 1977.

3 Pirsch, P. 'A new predictor for dpcm coding of tv signals', ICC '80 Conf., Seattle, Washington, June 1980, pp. 31.2.1 to 31.2.5.

4 Habibi, A. and Wintz, P. A., 'Image coding by linear transformation and block quantization', *IEEE Trans. on Communications Technology*, **COM-19**, no. 1, pp. 50–63, February 1971.

5 Ahmed, N., Natarajan, T. and Rao, K., 'On image processing and a discrete cosine transform', *IEEE Trans. on Computers*, **C-23**, no. 1, pp. 90–93, January 1974.

6 Davisson, L. D., 'Rate-distortion theory and application', *Proc. IEEE*, **60**, no. 7, pp. 800–08, July 1972.

7 Max, J., 'Quantizing for minimum distortion', *IRE Trans. on Information Theory*, **IT-16**, no. 1, pp. 7–12, March 1960.

8 Wong, W. C. and Steele, R., 'Adaptive coding of discrete cosine transform video telephone pictures', Picture Coding Symposium, Ipswich, England, July 1979, paper 12.3.

9 Haskell, B. G. and Steele, R., 'Audio and video bit-rate reduction', *Proc. IEEE*, **69**, no. 2, pp. 252–62, February 1981.

Manuscript first received by the Institution on 6th November 1980 and in revised form on 8th April 1981.

(Paper No. 2116/Comm 327)

UDC 621.396.665
Indexing Terms: Amplifiers, Control of gain

A review of automatic gain control theory

D. V. MERCY, B.Sc., C.Eng., M.I.E.E.*

SUMMARY
The paper presents the general theory of automatic gain control (a.g.c.). It includes discussions on loop gain, regulation and the relation between loop parameters and output level stability. An equivalent circuit is derived which is used in the study of dynamic behaviour. The work is extended to cover sampled data a.g.c. systems and design guidelines are derived for a variety of situations.

* *EMI Electronics Ltd., Hayes, Middlesex*

1 Introduction

The purpose of an automatic gain control (a.g.c.) circuit is to maintain the output level of an amplifier, for example, at an almost constant value, even though its input signal level may change substantially. Virtually all radio and television receivers use a.g.c. techniques so as to ensure a constant output level despite uncertain reception conditions. Other applications include the stabilization of signal generators and the maintenance of optimum signal levels on cable repeater systems. Similar techniques are used in certain types of power supply regulator.

The basic characteristic of an a.g.c. system is discussed first in Section 2 and methods of implementing it are discussed in general terms. There follows an analysis of the a.g.c. loop which demonstrates the importance of the control law of the variable gain element(s). To highlight this point further, a.g.c. behaviour is considered for several typical control laws.

System performance can be impaired by the presence of non-constant circuit elements and this is considered briefly (Sect 4).

A simple equivalent circuit of the a.g.c. loop is derived in Section 5 and is used as a basis for investigations of stability and transient performance later in the paper. In Section 8 the work is extended to gated (or sampled data) control systems.

Finally there are short discussions on transient behaviour during overload situations and on feedforward a.g.c.

2 Basic Concepts

Figure 1 shows the input–output characteristic of an a.g.c. system and, although somewhat idealized, it serves as a useful reference. For convenience, the scales of the axes are logarithmic.

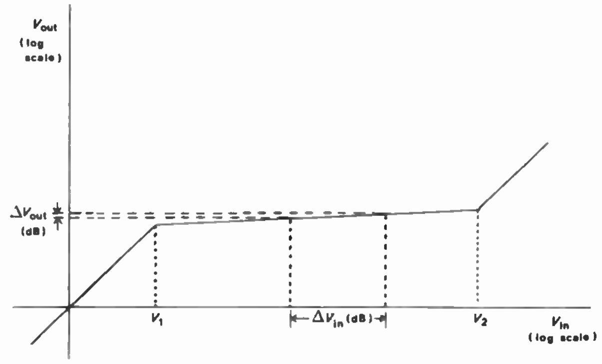


Fig. 1. A.g.c. characteristic

As indicated in the Figure, little or no a.g.c. action occurs when the input level is low and so any change in v_{in} results is an identical change in v_{out}.

When the input level reaches v_1, a.g.c. action begins and, for the range $v_1 \leqslant v_{in} \leqslant v_2$ the a.g.c. system attempts to maintain v_{out} constant in value. The measure by which the loop is successful has been variously called the 'figure of merit',[4] the 'compression ratio',[19] the

0033–7722/81/110579+12 $1.50/0

'regulation',[17] the 'flatness factor',[16] or the 'stiffness'[22] of the a.g.c. system. It can be defined thus:

$$\text{compression ratio} = M = \frac{\text{change in input level, expressed in dB}}{\text{change in output level, expressed in dB}}$$

As shown later, the value of M obtained from this equation is closely related to the loop gain of the system.

If v_{in} is further increased so that v_{in} exceeds v_2, the a.g.c. action ceases and once again the output level increases as shown in the Figure. Usually some form of current or voltage limiting (i.e. overloading) of one or more stages in the feedback loop determines the limits v_1 and v_2 of the effective a.g.c. range. It is possible for the upper limit to be set deliberately low to avoid instability problems and this point is discussed later.

An example of an a.g.c. loop, in schematic form, is given in Fig. 2. As shown, a level sensing amplifier controls two variable gain elements in the main amplifier chain in a manner that ensures the output level is maintained constant for a prescribed range of input levels.

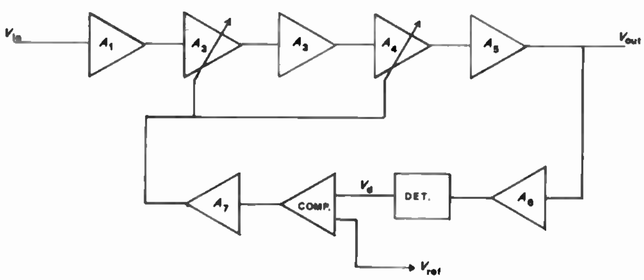


Fig. 2. Schematic of an a.g.c. system

Without additional information, the detailed behaviour of the loop cannot be determined. However, the signal level at which a.g.c. action begins can be estimated fairly easily. It occurs when $v_d \simeq v_{ref}$, which in this example is when $A_1 A'_2 A_3 A'_4 A_5 A_6 v_{in} \simeq v_{ref}$, where a lossless detector is assumed and A'_2 and A'_4 are the maximum gain values of A_2 and A_4 respectively.

There are many variations on the basic theme and some possible alternatives are:

(i) The main amplifier chain may be broadband or bandpass (as may the feedback amplifier A_6).

(ii) The gain controlled elements may differ greatly in type, number and circuit position according to system requirements. Important considerations are the dynamic range, operating frequency, tolerance to overloads, and the maintenance of good signal to noise ratio.[5,13]

(iii) The type of detector (or demodulator) is necessarily related to that property of the output signal to be controlled. For example, the signal characteristic of interest could be the peak-to-peak value of the r.f. signal, its mean value, the depth of modulation on a carrier, the power in one sideband, and so on.

(iv) The reference voltage, v_{ref}, may vary under external control. (If $v_{ref} = 0$ then the loop operates in the 'simple a.g.c.' mode, which is very common in inexpensive radio receivers).

(v) Further signal processing may be included in the loop following the detector or comparator.

It is beyond the scope of this paper to discuss the enormous variety of a.g.c. schemes but additional information can be found in the references listed. In particular, information on practical aspects of circuit design can be found in Crawford,[5] Rheinfelder,[20] Shirman[22] and Sutton.[23] Data on variable gain elements have been published by Harada,[6] Isaacson,[8] Kiehn,[9] Lea,[11] Lepoff[12] and Weaver.[30] The particular problems of television a.g.c. systems has been discussed by Clark,[4] Kiver,[10] Mills,[14] Nabe-yama[15] and Wendt.[31]

3 Preliminary Loop Analysis

Before proceeding with the analysis, it is possible to further simplify the circuit of Fig. 2 to that of Fig. 3.

In Fig. 3 the variable gain section P is assumed to have a nominal gain of unity, with all the forward gain of the loop lumped together as A. The detector is omitted from the circuit so that v_{in} and v_{out} are understood to refer specifically to the signal property under a.g.c. The net gain of all circuits following the comparator is absorbed into the value taken for the slope (dP/dv_c) of the variable gain section.

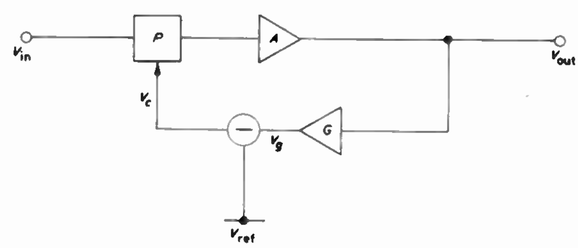


Fig. 3. Basic circuit for feedback a.g.c.

For convenience, the variable gain section is assumed to be voltage controlled, but in practice both current- and voltage-controlled circuits can be found.

At any given operating level, the input and output signals are, of course, related by $v_{out} = APv_{in}$. However, in the case of a.g.c. systems, fractional changes in the input and output levels are of more interest as given by equation (1):

$$\frac{\Delta v_{out}/v_{out}}{\Delta v_{in}/v_{in}} = \frac{1}{1 + AGv_{in}(dP/dv_c)} \qquad (1)$$

As the loop characteristic is non-linear, equation (1) is exact only if Δv_{in} is small.

In equation (1) the expression $AGv_{in}(dP/dv_c)$ represents the loop gain, L, of the system. Also, since

$$\frac{1}{P}\cdot\frac{dP}{dv_c} = \frac{2{\cdot}303}{20}\cdot K,$$

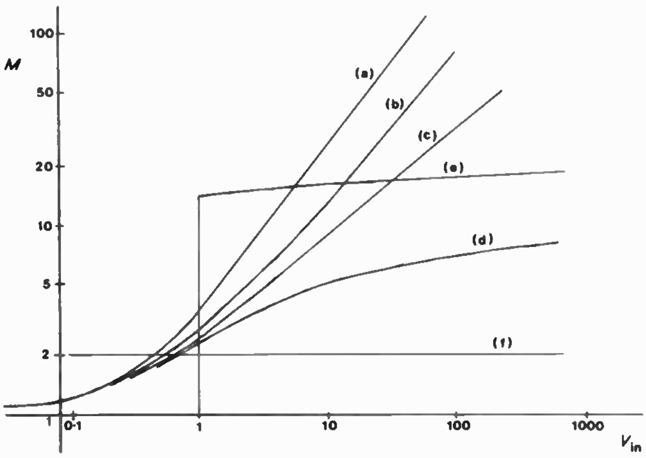


Fig. 4. Loop performance for various attenuator laws

where K is the slope of the voltage-controlled attenuator expressed in decibels per volt, the loop gain can be written as

$$L = AGv_{in}(dP/dv_c) = 0{\cdot}1151KGv_{out} = 0{\cdot}1151K(v_{ref}-v_c) \quad (2)$$

Also, since

$$\frac{\Delta v_{out}/v_{out}}{\Delta v_{in}/v_{in}} \simeq \frac{\Delta v_{out}(\text{dB})}{\Delta v_{in}(\text{dB})}$$

then

$$M = \frac{\Delta v_{in}(\text{dB})}{\Delta v_{out}(\text{dB})} = 1+L \quad (3)$$

Both P and K are functions of v_c (and hence functions of v_{in} and v_{out}) and in consequence the detailed behaviour of an a.g.c. system may not be obvious at first glance. However, on the occasions when the characteristics of P can be expressed in a convenient mathematical form, it is possible to derive an expression for M in terms of v_{in}, for example, by means of equations (2) and (3). In a more general case, provided details of K are known (whether by measurement or theory), it is possible to derive numerically the values of P, v_{in}, v_{out}, M, etc. by choosing the value of v_c as the starting point.

Invariably the law of the voltage- (or current-) controlled attenuator plays an important role in determining the value of the loop gain and its connection to signal levels within the system. This point is illustrated by the curves given in Fig. 4 in which M is plotted against v_{in} for several different attenuator characteristics. The curves are discussed below and further background material can be found in Appendix 1.

Curve (a) corresponds to a linear attenuator characteristic, i.e. $P \propto v_c$. In this case the loop gain is proportional to A, G and v_{in}. In theory very large loop gains are achievable, especially when signal levels are high. It is wise, therefore, to ensure that an upper limit is set to the a.g.c. range to avoid instability problems.

A somewhat similar characteristic results when the attenuator law is exponential, but with lower values of M, as shown by curve (b).

Curve (c) is for a square law characteristic and the relatively low values of M which occur when v_{in} is large (which is when v_c is small) are a consequence of the near-zero slope under these circumstances.

Curves (d) and (e) both correspond to the law $P(\text{dB}) \propto v_c$. In the case of curve (d), v_{ref} is set at zero (i.e. simple a.g.c.), and the regulation is seen to be poor. Little improvement is obtained by increasing the values of A or G because a situation of diminishing returns soon occurs. On the other hand, curve (e), which is for an identical attenuator characteristic, shows the benefit of deliberately inhibiting a.g.c. action when input levels are low.

Lastly, curve (f) is for the attenuator law $P \propto 1/v_c$. This is a case considered by Rheinfelder and M has the value 2 only, irrespective of signal level and the values of A and G.

In practice, four-quadrant multipliers have a linear law, as do signal attenuators of the balanced mixer type (when current driven). Balanced mixers (when voltage driven) have an exponential law. Some controlled attenuators have a straight line relationship between attenuation, in decibels, and the logarithm of the control voltage. These devices are satisfying, in fact, power law relationships of which 'linear' and 'square law' (approximately) are the most common. Finally, p-i-n diode attenuators (when appropriately driven[29]), dual gate f.e.t.s and current steering circuits all have characteristics which tend to satisfy the $P(\text{dB}) \propto v_c$ law.

It is clear from an inspection of equation (2) and Fig. 4 that the largest values of M occur when K increases rapidly in value as signal levels increase. Ultimately, once the dynamic range of the attenuator has been exceeded, a.g.c. action will cease and M will decrease towards unity, but this has not been shown in the Figure.

As an a.g.c. system has a non-linear characteristic, it is possible to express the relationship between v_{in} and v_{out} by means of a power series expansion and, in fact, this has been done for several of the examples in Appendix 1. Alternately, if an a.g.c. system is operating in a 'stable' condition, then changes in level away from this condition can be expressed by

$$v_{out}(\text{dBu}) = \Delta v_{out}(\text{dB}) + \overline{v_{out}}(\text{dBu})$$

i.e.

$$v_{out}(\text{dBu}) = \frac{\Delta v_{in}(\text{dB})}{\overline{M}} + \overline{v_{out}}(\text{dBu}) \quad (4)$$

where, for convenience, levels are expressed in dB with respect to unity (dBu), $\overline{v_{out}}(\text{dBu})$ is the mean (i.e. the desired) value of the output signal and $\overline{M}$ is given by, for example,

$$\overline{M} = 1 + 0{\cdot}1151\, K\, \overline{v_{out}} \quad (5)$$

Signal levels could have been taken with respect to 1 millivolt (dBmV), if preferred.

4 Sensitivity of the Output Level to Loop Parameter Changes

The main preoccupation of a.g.c. theory is with the relationship between input and output signal levels, as given by equation (4), for example. However, it is possible for the output to change for reasons other than changes in v_{in}. At worst, any intended benefits deriving from an a.g.c. system can be nullified by parameter changes within the loop.

It is possible to obtain results either by partial differentiation of loop equations or by a direct consideration of any circuit change on the output level.

Changes in A or P are not very serious, if the loop gain is high, because their effects on the output are reduced by a factor $1/(1+L)$ times. On the other hand, any changes in G or v_{ref} (expressed in dB), results in an almost identical change in v_{out}.

It is important, therefore, to ensure that all components which make up G (amplifiers, filters, detector gain, etc.) and all parameters which influence v_{ref} (i.e. the supply for v_{ref}, the offset voltage of the comparator, the detector d.c. voltage, etc.) are all time and temperature stable and insensitive to supply voltage variations.

5 An Equivalent Circuit of the A.G.C. Loop

It is useful to have an equivalent circuit of the a.g.c. loop which is directly comparable with the standard feedback amplifier circuit, as this allows the well-known results of feedback theory to be applied directly to the a.g.c. loop, particularly those relating to dynamic behaviour.

Oliver[17] in an early paper subdivided an a.g.c. circuit into 'α' and 'β' portions by inspection and proceeded with the analysis from that viewpoint. The method used here is based on that of Victor and Brockman,[28] but differs in some respects. The procedure is summarized below and given in detail in Appendix 2.

Firstly, the voltage-controlled attenuator of Fig. 3 is replaced by two elements, a 'divider circuit' and a 'divider controlling amplifier' (DCA), as shown in Fig. 5(a) where the slope of the DCA, in dB per volt, is given by

$$K = \frac{d}{dv_c} 20 \log_{10} v_y \tag{6}$$

Corresponding to a given control voltage v_c, the DCA produces an output v_y such that, as v_{in} varies, the loop acts to ensure that v_{in}/v_y remains as constant as possible.

If signal level changes are not too great, the control voltage output from the comparator can be expressed in logarithmic terms as

$$v_c = k.20\log_{10} \frac{v_{in}}{v_y \overline{v_{in}}} \tag{7}$$

where

$$k = 0{\cdot}1151\, AG\overline{v_{in}} = 0{\cdot}1151\, v_{ref} \tag{8}$$

By combining equations (6) and (7) and noting that

$$\Delta v_{in}(\text{dB}) = v_{in}(\text{dBu}) - \overline{v_{in}}(\text{dBu})$$

then

$$\Delta v_{in}(\text{dB}) = \frac{1+kK}{kK} v_y(\text{dBu})$$

$$\frac{v_c(\text{volts})}{\Delta v_{in}(\text{dB})} = \frac{k}{1+kK}$$

and

$$\frac{\Delta v_{in}(\text{dB})}{\Delta v_{out}(\text{dB})} = M = 1+kK$$

Figure 5(b) gives the block diagram which corresponds to these equations. Note that $kK = L$, the loop gain.

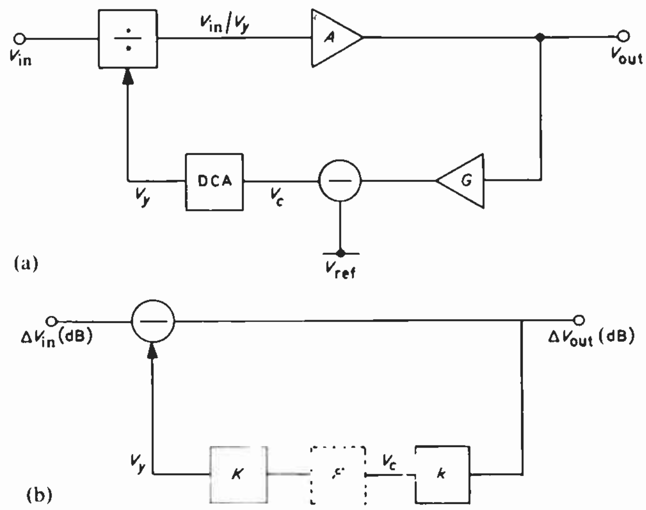


Fig. 5. Equivalent circuit (c) of a.g.c. loop

It can be seen that the equations closely resemble those of feedback theory and, in particular, the term Δv_{out} corresponds to the 'error voltage' which is output from the summation point of a feedback loop.

The equivalent circuit in Fig. 5(b) is peculiar in that parameters are expressed in mixed units, i.e. Δv_{in} and Δv_{out} are in dB, v_y is in dBu, v_c is in volts, k in volts per dB and K in dB per volt. Also, it does not give the output level in absolute terms, but this can be derived easily by means of equation (4).

When filter sections are included in the loop, it is convenient to lump them all together as 'F' and to keep k and K frequency independent. Narrow band r.f. filters in the loop can be replaced by low-frequency analogues and included in F. In Fig. 5(b) the location of F is shown dotted.

A more general expression for M now becomes

$$M = 1+FkK = 1+FL$$

6 Dynamic Characteristics and Stability

The roots of $(1+FL)$ determine the small signal characteristics of an a.g.c. loop in the same way that the roots of $(1+\mu\beta)$ do for a conventional feedback amplifier. The standard results may be found by a consideration of root locus plots in the 'p' plane.

When F is a single RC time-constant, i.e. $F = 1/(1+\tau_1 p)$, then the loop is always stable regardless of loop gain. When the loop is subjected to a unit step change, the output change is given by

$$\Delta v_{out}(p) = \frac{(p+1/\tau_1)}{p\left(p+\frac{1+L}{\tau_1}\right)}\ \text{dB} \qquad (9)$$

and

$$\Delta v_{out}(t) = \left(\frac{1}{1+L}+\frac{L}{1+L}\cdot\exp-[(1+L)t/\tau_1]\right)\text{dB}$$

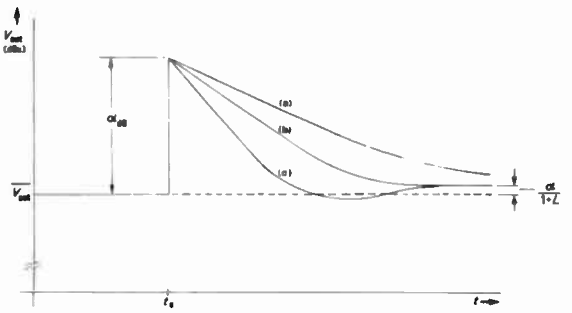


Fig. 6. Transient behaviour of an a.g.c. loop
(a) Single time-constant; (b) Two time-constants with critical damping; (c) Two time-constants, Butterworth characteristic

A typical response with $L = 9$ is given in Fig. 6, curve (a).

Virtually all previous papers assume only a single time-constant in the loop. This is, in fact, a desirable objective, as shown later, and is very often a valid assumption. However, the existence of other time-constants in the loop should not be overlooked because they can have an adverse effect on the transient response and stability.

A few guidelines for loops containing two or more time-constants are given below. It is assumed that

$$\tau_1 \gg \tau_2 \gg \tau_3 \gg \sum_4^N \tau_n$$

(where the 'very much greater than' sign implies, roughly, an order of magnitude difference).

Then for

(i) critical damping $L \simeq \tau_1/4\tau_2$

(ii) Butterworth $L \simeq \tau_1/2\tau_2$

(iii) instability (a) with 1 or 2 poles—no limit on L
(b) with 3 or more poles—$L \simeq \tau_1/\tau_3$.

When pole cancellation techniques have been used in the loop (i.e. 'modified lag', 'lead' or 'lead-lag' compensation,[2,17] for example), then τ_1, τ_2 etc. refer to the time-constants that remain after the compensation procedure has been carried out.

Typical transient responses for a step input are given in Figs 6(b) and 6(c) for a critically damped loop (poles due to τ_1 and τ_2 coincident and real) and a Butterworth situation respectively. As in Fig. 6(a), $L = 9$.

7 A.G.C. Systems in Cascade

In several respects, an a.g.c. system exhibits the same properties as a high-pass filter. It will act on a slowly-varying input signal so that at the output the low-frequency component is virtually absent. On the other hand, if the input signal varies rapidly, so that the a.g.c. loop is unable to respond, then the high-frequency components appear at the output unimpaired. In fact, the transient responses of Fig. 6 are consistent with a high-pass characteristic.

Now it has been shown that if a step transient is applied to a cascade of high-pass filters, the output exhibits overshoots and ringing. Valley and Wallman,[27] for example, considered a cascade of single-pole filters and some results are given in Fig. 2.17 of their book. Similar results occur when a.g.c. systems are cascaded, a situation which frequently arises in cable repeater installations.

In the case of a cascade of n repeater amplifiers, where the gain of one amplifier equals the cable loss per section, all amplifiers have the same output level under stable conditions. When an input change occurs, Δv_{in}(dB), the final output signal is given by

$$v_{out}(\text{dBu}) = \frac{\Delta v_{in}(\text{dB})}{M^n} + \overline{v_{out}}(\text{dBu})$$

If each loop contains a single effective time-constant, τ_1, then for a unit step input, the change at the output is

$$\Delta v_{out}(p) = \frac{(p+1/\tau_1)^n}{p\left(p+\frac{1+L}{\tau_1}\right)^n}\ \text{dB}$$

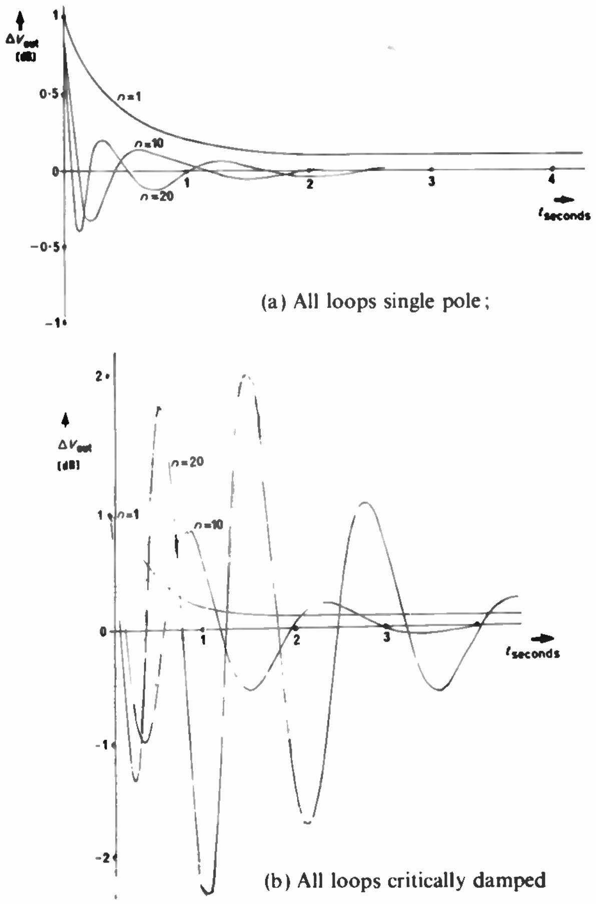
Providing L is large, the results approximate to those discussed by Valley and Wallman. Additional results, for large values of n, have been presented by Chisholm.[3]

When more than one active pole is present in each a.g.c. loop the transient responses show increased amounts of overshoot and ringing. Figures 7(a) and 7(b) show transient responses for cascades of single-pole and critically-damped loops respectively.

8 Sampled Data A.G.C. Systems

It has been assumed, so far, that the output signal of an a.g.c. system is monitored continuously by the feedback loop. However, many systems operate by examining the output level only during small, discrete, regular intervals of time. For the remainder of the time, the control loop is effectively open-circuit. These systems belong to the family of sampled data control systems.[18,25] Television receivers with gated a.g.c. operate in this way.[4]

In order to determine conditions for an overshoot-free transient response and for stability, z-transform analysis provides a useful tool. For a sampled data system, all poles of the z-transform must lie within the unit circle to ensure stability and for an overshoot free transient



(a) All loops single pole;

(b) All loops critically damped

Fig. 7. Transient responses for a.g.c. loops in cascade

response all poles must be real and lie within the range 0 to +1.

In the simplest approach to sampling systems (mathematically speaking) the sampling time is considered to be very small and the output of the sampler is an impulse of 'strength' equal to the instantaneous amplitude of the sampled waveform. Provided the sampler is immediately followed by a zero-order hold (clamp) circuit and the following listed conditions are satisfied, then this idealized approach successfully represents the behaviour of a practical a.g.c. circuit:

(i) The hold capacitor attains virtually the full instantaneous value of the voltage being sampled.

(ii) The sampling duration is very short so that the sampled voltage changes very little during the 'on' time.

(iii) The hold capacitor retains its voltage with negligible decay during the 'off' period of the sampler.

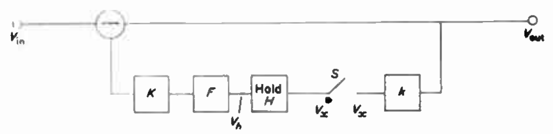


Fig. 8. A.g.c. system with impulse sampler

In Section 8.2 some results are presented which are based on the impulse sampler approach. A discussion of 'finite-width' samplers is given in Section 8.3.

8.1 Sample-and-hold Circuit

Figure 8 shows the equivalent circuit of an a.g.c. loop which contains a sample-and-hold circuit. The input and output levels are related by

$$v_{out}(z) = \frac{v_{in}(z)}{1 + L \cdot HF(z)}$$

where $L = kK$ and $HF(z)$ is the z-transform of $H(p) \cdot F(p)$. To ascertain the dynamic behaviour of the loop it is necessary to investigate the roots of $1 + L \cdot HF(z)$. A summary of results is given below and further details can be found in Appendix 3. In all cases, H is assumed to be a zero-order hold and T is the time period of the sampler.

(i) $F = 1$

No stability problems and no condition for critical damping.

(ii) $F = 1/(1+\tau_1 p)$

Critical damping occurs when $L \simeq \tau_1/T$. Instability is when $L \simeq 2\tau_1/T$.

(iii) $F = 1/(1+\tau_1 p)(1+\tau_2 p)$

For critical damping, results are best summarized in tabular form. Note that Z_d is the position of the coincident (real axis) poles and L has an approximate value of 10 or greater.

Table 1

T/τ_2	$L\tau_2/\tau_1$	LT/τ_1	Z_d
small	1/4	small	1
1	1/5·1	1/5·1	0·65
10	1/17·3	1/1·7	0·23
	0	1	0

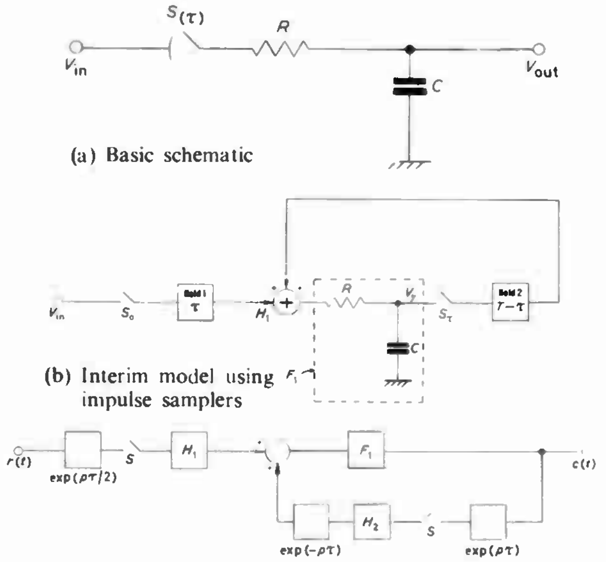
It can be seen that to maintain critical damping as τ_2 is increased in value (with L unchanged), it is necessary to reduce the value of T/τ_1. The fact that Z_d increases towards unity indicates that longer settling times are an undesirable consequence.

Conditions for stability cannot be presented in such a detailed way. This is because so many conditions are possible, depending on the relative values of T, τ_1 and τ_2.

However, if the condition $L < 2\tau_1/T$ is satisfied, then stability is assured. It should be borne in mind that under certain circumstances (e.g. when $\tau_1 \ll T$) this inequality gives somewhat pessimistic values for L.

8.2 Finite-width Sampler

In a practical situation the feedback loop operates with sampling pulses of finite duration and the assumptions



(a) Basic schematic

(b) Interim model using impulse samplers

(c) Final model including pseudo-time-advance and -delay circuits

Fig. 9. Finite width sampler

listed at the end of Section 8.1 may not be valid. Techniques[26] have been developed to deal with the situation, but, unless numerical data are available from the outset, they are cumbersome to use. A simplified model of a finite width sampler is used here and it is shown in Fig. 9.

A more detailed description of the circuit is given in Appendix 3, but basically the 'actual' sample and hold shown in Fig. 9(a) is represented by a loop containing two impulse samplers, each with a hold circuit, as shown in Figs. 9(b) and 9(c). In Fig. 9(c), artificial time advances and delays are included in the loop to allow the impulse samplers to operate in synchronism as required by z-transform theory. In effect, the first sampler takes the value of the input at a time halfway through the sampling pulse interval and this is held for τ seconds by H_1. The second sampler takes the value of the output at the end of τ seconds and this is held by H_2 for $T-\tau$ seconds.

In Fig. 10 the sample and hold model is shown included in an a.g.c. circuit. An analysis of the circuit

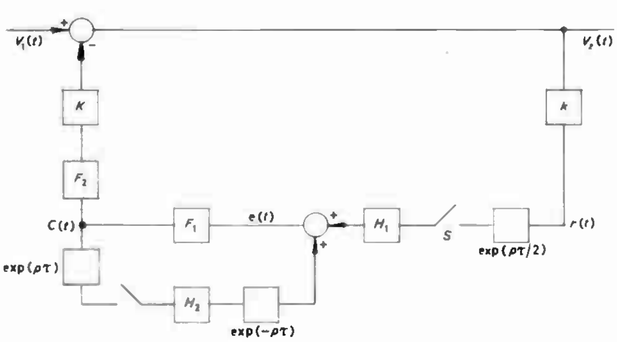


Fig. 10. A.g.c. system including S and H model

gives

$$\frac{v_{out}}{v_{in}} = \frac{1}{1+L\left\{F_1F_2H_1(z,\tau/2)+\dfrac{F_1F_2H_2(z,-\tau/2)\cdot F_1H_1(z,\tau)}{1-F_1H_2(z)}\right\}} \tag{10}$$

As in Section 8.2, the behaviour of the poles in the z-plane is of interest. If H_1 and H_2 are zero-order holds and the filters are single-pole networks with $F_1 = 1/(1+\tau_1 p)$ and $F_2 = 1/(1+\tau_2 p)$, then the denominator of equation (10) reduces to a quadratic in z. The results of the tests for critical damping and instability carried out on this quadratic are summarized in Tables 2 and 3 respectively.

Table 2

Values for L for critical damping

	T/τ_2 small			T/τ_2 large
	T/τ_1 small	T/τ_1 large, $\alpha T/\tau_1$ small	T/τ_1 large, $\alpha T/\tau_1$ large	
$\tau_1 \gg \tau_2$	$\tau_1/4\alpha\tau_2$	—	—	$\tau_1/\alpha T$
$\tau_2 \gg \tau_1$	$\tau_1/4\alpha\tau_2$ if $\tau_1 \gg \alpha\tau_2$; $\alpha\tau_2/4\tau_1$ if $\alpha\tau_2 \gg \tau_1$		τ_2/T	Model not valid, see note (v)
$\tau_1 = \tau_2$	$1/4\alpha$	$\tau_1/\alpha T$	—	

Table 3

Values of L for instability

	T/τ_2 small			T/τ_2 large
	T/τ_1 small	T/τ_1 large, $\alpha T/\tau_1$ small	T/τ_1 large, $\alpha T/\tau_1$ large	
$\tau_1 \gg \tau_2$	$4\tau_1/\alpha T\cdot\tau_2/T$	—	—	$2\tau_1/\alpha T$
$\tau_2 \gg \tau_1$	$4\tau_1/\alpha T\cdot\tau_2/T$		$2\tau_2/T$	no limit (always stable)
$\tau_1 = \tau_2$	$4\tau_1/\alpha T\cdot\tau_2/T$	$2\tau_1/\alpha T$	—	—

To assist in interpreting the tables, note that 'large' means greater than 3 (very approximately) and 'small' means less than 0·4 (very approximately). Also $\tau = \alpha T$.

Some general points can be made from an inspection of the tables, as follows.

(i) Regardless of the value of L, a loop can always be made stable by choosing a value for α that is sufficiently small.

(ii) Similarly, stability is assured by choosing a sufficiently high sampling frequency (i.e. a small value of T).

(iii) When the value of $\alpha T/\tau_1$ is large, the sampler approximates to an impulse sampler and the results are identical to those of Section 8.1 (ii).

(iv) When both T/τ_1 and T/τ_2 are small (i.e. the sampling period is short compared to loop time-constants), the result of the sampler is, effectively, to increase the time-constant of the first filter by $1/\alpha$ times. Thus the conditions for critical damping and stability can be obtained directly from the results given in Section 5.

(v) If both $\alpha T/\tau_1$ and T/τ_2 are very large, it is possible for the loop to settle in time τ seconds (i.e. during the time of a single sampling pulse). In consequence, the relationship between τ_1 and τ_2 is the same as the continuous case discussed in Section 5.

9 Loop Performance during Overloads

Nolle[16] has discussed overload effects in a.g.c. systems. He considered, in particular, loops which contained only a single time-constant (although the arguments can be extended to more complex situations).

When in overload the effective reduction in the value of the time-constant (by $1/(1+L)$ times) does not occur and voltages decay according to the 'natural' value of the time-constant. Nolle showed that the actual rate of gain change of an a.g.c. system, when in overload, depends on the instantaneous voltage existing across the a.g.c. capacitor and bears no relation to the magnitude of the change.

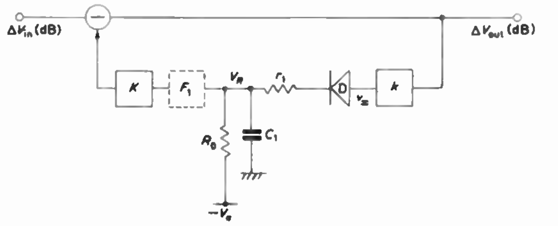


Fig. 11. Control loop with peak rectifier

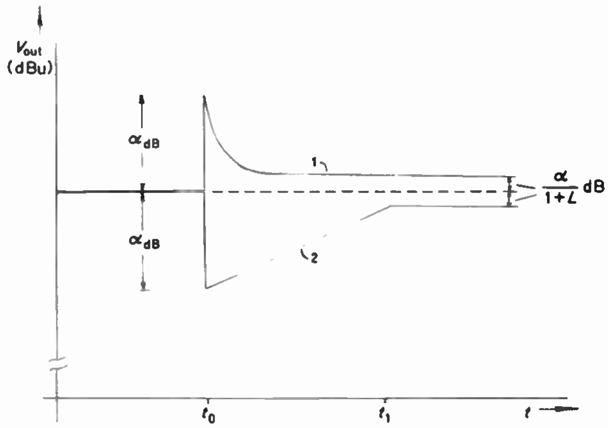


Fig. 12. Transient responses for circuit with a peak rectifier

This behaviour is particularly relevant to an a.g.c. loop in which a peak rectifier circuit has been incorporated. Such a loop is prone to overload when subject to sudden changes in input level. For example, the system shown in Fig. 11, which contains a positive peak rectifier, overloads immediately when the input change is a negative step. Figure 12 gives typical responses for this circuit when input steps are positive or negative (with $F = 1$).

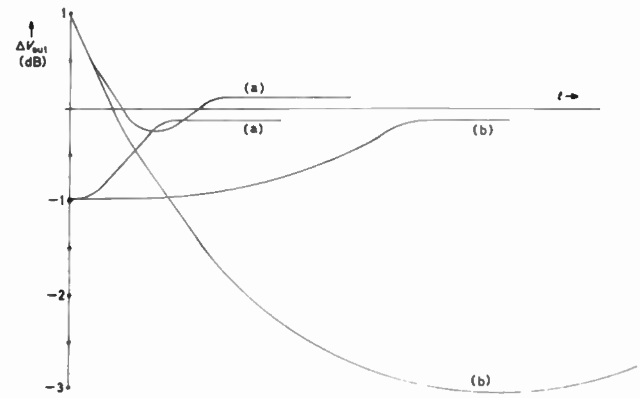


Fig. 13. Responses of a.g.c. loop with peak rectifier and additional filter
(a) R_0C_1/V_A small; (b) R_0C_1/V_A large

When there are other time-constants in the loop, in addition to the peak rectifier circuit, the transient behaviour becomes very complicated and, in fact, severe overshoots can occur if time-constant values are injudiciously chosen. This is demonstrated in Fig. 13. The filter F is given by $F = 1/(1+\tau_1 p)$ and transient responses are given for different values of the peak rectifier time-constant R_0C_1. As shown, unsatisfactory overshoots occur if R_0C_1 is too large.

Additional problems also arise when a.g.c. loops are cascaded and the overshoots and ringing that occur are several times larger than the corresponding results with no peak rectifiers present.

Of course, there may be a good reason to include a peak rectifier in the a.g.c. loop,[1,3,21] for example,

(i) the peak value of the output signal is to be maintained constant,

(ii) a fast attack, slow decay response can be desirable in some circumstances,

(iii) the use of a peak rectifier reduces 'transfer modulation effects'.

However, it is clear that more care is needed in the choice of loop time-constants than in cases previously discussed.

10 Feedforward A.G.C.

In all the previous Sections a feedback a.g.c. loop has been assumed, but there has recently been renewed interest in feedforward systems.[19]

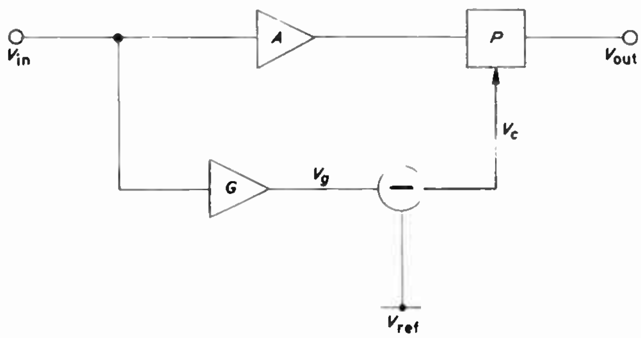


Fig. 14. Schematic for feedforward a.g.c.

Figure 14 gives a simplified schematic of a feedforward loop. With $v_{ref} = 0$,

$$M = \left(1 + \frac{1}{P}\cdot\frac{dP}{dv_c}\right)^{-1}$$

If the law of the voltage controlled attenuator is given by $P \propto 1/v_c$, then $M = \infty$ and the loop exhibits perfect regulation. However, as it is unlikely that the required characteristic can be maintained over a wide range of signal levels, it is advisable to use, in addition, a feedback a.g.c. circuit to provide some initial stabilization.[21]

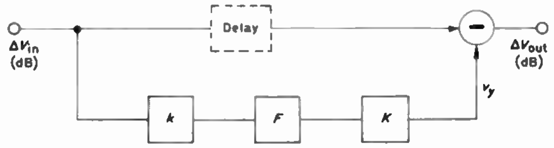


Fig. 15. Equivalent circuit for feedforward a.g.c.

A small-signal equivalent circuit can be derived for the feedforward loop as shown in Fig. 15, with the values of k and K given by equations (6) and (8) and the filter section F included for generality. The circuit gives

$$\frac{\Delta v_{out}}{\Delta v_{in}} = \frac{1}{M} = 1 - FkK$$

No stability problems arise because there in no feedback loop, i.e. the circuit is 'open loop'. However, unacceptable transient responses can occur if circuit values are poorly chosen. Also, it is necessary for the two signal paths to have the same time delay at all frequencies of interest, if the full benefits of the circuit are to be realized.[19] For example, the speed of response of a forward acting a.g.c. loop can be much higher than is possible with a feedback system.[7,19]

11 Conclusions

The basic theory of a.g.c. systems has been considered. In general the operating signal level and the characteristics of the variable gain elements play a significant role in determining the loop gain of a system.

For studies of stability, the well known rules of feedback theory have been shown to apply with only minor differences.

In the case of sampled data (gated) a.g.c., the time period of the sampler imposes additional restraints on the permissible loop gain and some results have been tabulated.

12 References

1 Arnold, B., 'Agc/asc design', *IEEE Trans. on Cable Television*, CATV-3, pp. 137–40, 1978.
2 Barna, A., 'Operational Amplifiers', pp. 85–102 (Wiley, New York 1971).
3 Chisholm, E. W., 'Considerations on the application of agc/asc amplifiers in catv systems', *IEEE Trans.*, CATV-1, pp. 40–62, 1976.
4 Clark, E. G., 'Automatic gain control' in 'Television Engineering', Ed. D. Fink, pp. 16.142–16.148, (McGraw Hill, New York, 1957).
5 Crawford, J. S., 'A review of automatic gain control in transistor television receivers', *Proc. IREE* (Aust.), **27**, pp. 1–9, January 1966.
6 Harada, Y., 'A precision a.g.c. circuit using analogue computing ics', *Electronic Engineering*, **52**, no. 641, p. 19, August 1980.
7 Hopper, A. L., 'Forward acting gain control', *Proc. Nat. Electronics Conf.* **19**, 1963, pp. 358–70.
8 Isaacson, I., Moody, G., Jackson, I., and Curtis, B., 'Vhf and uhf amplifiers for cable t.v.', *SCTE Cable T.V. Engng*, **2**, no. 1, pp. 47–57, Dec. 1977 and April 1978, **2**, no. 2, pp. 108–22.
9 Kiehn, H. C., 'Application of the CA3028 IC RF amplifier', RCA Application Report ICAN-5337, 1967.
10 Kiver, M. S., 'Television Simplified', pp. 160–79 (Van Nostrand, New York, 1962).
11 Lea, R. M., 'The junction fet as a voltage controlled resistance', *Wireless World*, **78**, pp. 337–40, July 1972, and **78**, pp. 394–6, August 1972.
12 Lepoff, J. H., 'A new pin diode for uhf–vhf applications', *IEEE Trans. on Broadcast and Television Receivers*, BTR-**17**, pp. 10–15, 1971.
13 McFee, R. and Dick, S., 'Reducing non-linear distortion due to agc circuits', *Electronics Letters*, **1**, no. 4, pp. 102–4, June 1965.
14 Mills, T. B. and Suzuki, H. S., 'Design concepts for low-cost transistor AGC systems', *IEEE Trans.*, **BTR-14**, pp. 38–43, October 1968.
15 Nabe-Yama, H. and Miyazaki, G., 'An agc system design based upon the d.c. restoration and the dynamic characteristic', *IEEE Trans.*, **BTR-16**, pp. 329–38, 1970.
16 Nolle, A. W., 'Adjustment speed of automatic volume control systems', *Proc. IRE*, **36**, pp. 911–16, July 1948.
17 Oliver, B. M., 'Automatic volume control as a feedback problem', *Proc. IRE*, **36**, pp. 466–73, April 1948.
18 Ragazzini, J. R. and Franklin, G. F., 'Sampled Data Control Systems' (McGraw-Hill, New York, 1958)
19 Rawlings, A. J., McGeehan, J. P. and Gosling, W., 'Forward feeding agc with extended signal delays', Conference on Radio Receivers and Associated Systems, July 1978, pp. 85–92, IERE Conf. Proc. No. 40.
20 Rheinfelder, W., 'Designing automatic gain control for transistorised amplifiers', *Electronic Equip. Eng.*, December 1964, pp. 43–7 and January 1965, pp. 53–7.
21 Rheinfelder, W., CATV System Engineering', (Tab, New York, 1966, Foulsham-Tab, New York, 1976).
22 Shirman, J., 'Designing a stable transistor agc amplifier', *Electronic Design*, May 1978, pp. 19–23.
23 Sutton, R. T., 'Designing AGC systems', *Electronics Industry*, May 1978, pp. 19–23.
24 Thaler, G. J. and Brown, R. G., 'Servomechanism Analysis', pp. 192–237, (McGraw-Hill, New York, 1953).
25 Tou, J. T., 'Sampled Data Control Systems', (McGraw-Hill, New York, 1959).
26 ibid., pp. 562–87.
27 Valley, G. E. and Wallman, H., 'Vacuum Tube Amplifiers' (McGraw-Hill, New York, 1948 and Dover, New York, 1965).
28 Victor, W. K. and Brockman, M. H., 'The application of linear servo theory to the design of agc loops', *Proc. IRE*, **48**, pp. 234–8, February 1960.
29 Watkins-Johnson Co., Design Data for Attenuator Type WJ-G1 and Linearizer Type WJ-LG1
30 Weaver, S., 'Dual-gate mosfets in tv if amplifiers', *IEEE Trans.*, **BTR-16**, pp. 96–110, May 1970.
31 Wendt, K. R. and Schroeder, A. C., 'Automatic gain controls for television receivers', *RCA Rev.*, **9**, no. 3, pp. 373–93, September 1948.

13 Appendix 1: Attenuator Characteristics

(i) If the attenuator characteristic is a linear function of the control voltage, then $P = Tv_c$ where T is the slope of the device in volts^{-1}. In this case

$$M = 1 + AGTv_{in}$$

and the input–output relationship is

$$v_{in} = \frac{1}{TA}\frac{v_{out}}{(v_{ref} - Gv_{out})}$$

or

$$v_{out} = \frac{TAv_{ref} \cdot v_{in}}{1 + TAGv_{in}}$$

(ii) When the attenuator characteristic is exponential, i.e. $P = D[\exp(Ev_c) - 1]$, where D and E are constants, then

$$M = 1 + GEv_{in} + GEADv_{out} = 1 + GEADv_{in}\exp(Ev_c)$$

and

$$v_{in} = \frac{v_{out}\exp(EGv_{out})}{DA[\exp(EGv_{ref}) - \exp(EGv_{out})]}$$

(iii) When $P = Dv_c^2$, where D is a constant, then

$$M = 1 + \frac{2Gv_{out}}{v_c} = 1 + \frac{2Gv_{out}}{v_{ref} - Gv_{out}}$$

and

$$v_{in} = \frac{v_{out}}{AD(v_{ref} - Gv_{out})^2}$$

(iv) If P, expressed in dB, is a linear function of the control voltage, then $K = K_0$, a constant.

(a) $v_{ref} = 0$

In this case

$$M = 1 + K_0 Gv_{out} = 1 - 0{\cdot}1151 P_{dB}$$

and

$$v_{in} = v_{out}\ \exp(Jv_{out}) \simeq v_{out} + Jv_{out}^2 + J^2 v_{out}^2 + \ldots$$

where

$$J = 0{\cdot}1151 K_0 G$$

(b) $v_{ref} \neq 0$

If when $v_g = v_{ref}$ (in Fig. 3), we define v_{in}, v_{out} and M to have their 'mean' values, $\overline{v_{in}}$, $\overline{v_{out}}$ and $\overline{M}$, respectively, and put $P = 1$, then

$$\overline{M} = 1 + 0{\cdot}1151 K_0 v_{ref} = 1 + 0{\cdot}1151 K_0 AG\overline{v_{in}}$$

it is then possible to derive

$$M = \overline{M} - 0{\cdot}1151 P_{dB} \quad \text{and} \quad v_{in} = \frac{M-1}{\overline{M}-1}\overline{v_{in}}$$

The input and output levels are related by

$$v_{in} = Hv_{out}\exp(Jv_{out}) \approx H(v_{out} + Jv_{out}^2 + \ldots)$$

where

$$H = \frac{\exp(-0{\cdot}1151 K_0 v_{ref})}{A}$$

(v) When $P = D/v_c^m$, where D is a constant, and $v_{ref} = 0$ (i.e. simple a.g.c.), then

$$M = 1 + m$$

which is independent of A, G, D and signal levels.

Also

$$v_{in} = \frac{G^m}{AD}(v_{out})^{1+m}$$

To enable meaningful comparisons to be made, all attenuator characteristics were normalized so that the value of P varied from unity to zero (approx.) for a control voltage range of 10 volts. The laws used in the derivation of Fig. 4 were as follows:

$$P = 0{\cdot}1\ v_c$$
$$P = 0{\cdot}156[\exp(0{\cdot}2v_c) - 1]$$
$$P = 0{\cdot}01 v_c^2$$
$$P = 10^{0{\cdot}2v_c}\ \text{(i.e. } K_0 = 4\ \text{dB/volt)}$$

14 Appendix 2: Equivalent Circuits

Consider the schematic of Fig. 5(a). The DCA produces an output voltage v_y for a given input voltage v_c. When the loop is in its 'mean' condition $v_c = 0$ and $v_y = 1$ so that $AG\overline{v_{in}} = v_{ref}$.

In practice, for values of v_{in} other than the mean, an error is detected and the output of the comparator circuit v_c will be given by

$$v_c = AG\frac{v_{in}}{v_y} - v_{ref} = AG\overline{v_{in}}\left\{\frac{v_{in}}{v_y \overline{v_{in}}} - 1\right\}$$

By using the expansion of $\log_e(1+x)$ this can be written

$$v_c = k\ 20\log_{10}\frac{v_{in}}{v_y \cdot \overline{v_{in}}}$$

where

$$k = \frac{2{\cdot}303}{20}AG\overline{v_{in}} = \frac{2{\cdot}303}{20}v_{ref}$$

If the sensitivity of the DCA, expressed in dB per volt, is K, then

$$\frac{d}{dv_c}20\log_{10} v_y = K$$

Integrating both sides and noting that $v_y = 1$ when $v_c = 0$, gives

$$20\log_{10} v_y = Kv_c$$

Hence

$$20\log_{10} v_y = Kk\ 20\log_{10}\frac{v_{in}}{v_y \cdot \overline{v_{in}}}$$

For convenience, express in dB with respect to unity, so that the equation becomes

$$v_y(\text{dBu}) = Kk\{v_{in}(\text{dBu}) - v_y(\text{dBu}) - \overline{v_{in}}(\text{dBu})\}$$

Rearranging and noting that $\Delta v_{in}(\text{dB}) = v_{in}(\text{dBu}) - \overline{v_{in}}(\text{dBu})$, then

$$\Delta v_{in}(\text{dB}) = \frac{1+kK}{kK}v_y(\text{dBu})$$

Also, since

$$\Delta v_{out}(\text{dB}) = v_{out}(\text{dBu}) - A\overline{v_{in}}(\text{dBu})$$
$$= v_{in}(\text{dBu}) - v_y(\text{dBu}) - \overline{v_{in}}(\text{dBu})$$

then

$$\Delta v_{out}(\text{dB}) = \frac{v_y(\text{dBu})}{kK}$$

and so

$$\frac{\Delta v_{out}(\text{dB})}{\Delta v_{in}(\text{dB})} = \frac{1}{M} = \frac{1}{1+kK}$$

15 Appendix 3: Sampled Data A.G.C. Systems

15.1 Sample-and-hold Circuit (Sect. 8.1)

(i) $F = 1$ (i.e. no filter)
In this case

$$\Delta v_{out}(z) = \frac{\Delta v_{in}(z)}{1+L}$$

(ii) $F = 1/(1+\tau_i p)$

Here

$$\Delta v_{out}(z) = \frac{[1-\exp(-T/\tau_1)z^{-1}]\Delta v_{in}(z)}{1+z^{-1}[L-L\exp(-T/\tau_1)-\exp(-T/\tau_1)]}$$

(iii) $F = 1/(1+\tau_1 p)(1+\tau_2 p)$ with $\tau_1 \gg \tau_2$

In this case

$$\Delta v_{out}(z) = \frac{(1-Rz^{-1})(1-Sz^{-1})\cdot\Delta v_{in}(z)}{1+z^{-1}\left(-R-S+L+\frac{LR\tau_1}{\tau_1-\tau_2}-\frac{LS\tau_2}{\tau_2-\tau_1}\right)+ +z^{-2}\left(RS+LRS+\frac{LS\tau_1}{\tau_2-\tau_1}+\frac{LR\tau_2}{\tau_1-\tau_2}\right)}$$

where $R = \exp(-T/\tau_1)$ and $S = \exp(-T/\tau_2)$

Critical damping occurs when T/τ_1 is small ($< 0{\cdot}1$) and τ_1/τ_2 is large (> 10), provided that L is reasonably large ($\geqslant 10$). Under these circumstances it is possible to write,

$$L\frac{\tau_2}{\tau_1} \simeq \frac{1-S}{\left(\sqrt{\frac{T}{\tau_2}}+\sqrt{1-S}\right)^2}$$

and the double pole is situated at z_d, where

$$z_d \simeq \frac{1+S}{2} - \frac{1-S}{2}\cdot\frac{\left(\sqrt{\frac{T}{\tau_2}}-\sqrt{1-S}\right)}{\left(\sqrt{\frac{T}{\tau_2}}+\sqrt{1-S}\right)}$$

For stability, it is necessary to ensure that the poles do not leave the unit circle. Now with a denominator of the form $A+Bz^{-1}+Cz^{-2}$, both roots will remain within the unit circle provided that the following three conditions are satisfied.

$$C < A; \quad B < A+C \quad \text{and} \quad -B < A+C$$

As stated in Section 8, according to the values of L, T/τ_1 and T/τ_2, several different conditions for stability are obtained. However, the inequality $L < 2\tau_1/T$ covers all situations.

15.2 Finite-width Sampler (Sect 8.2)

An analysis of the circuit in Fig. 9(c) gives as the result

$$C(z,\tau/2) = R(z,\tau/2)\left[H_1F_1(z,\tau/2)+ + \frac{H_2F_1(z,-\tau/2)\cdot H_1F_1(z,\tau)}{1-H_2F_1(z)}\right]$$

where $C(z,\tau/2)$ and $R(z,\tau/2)$ are the Z-transforms of $C(p)$ and $R(p)$, but corresponding to time instants occurring at the halfway points of the sampling pulses. For example,

$$C(z,\tau/2) \equiv Z\{C(p)\cdot\exp(p\tau/2\}$$

The analysis of an a.g.c. loop which includes the sample-and-hold model is given in equation (10) of the paper.

Now if we put $\tau = \alpha T$, with $0 \leqslant \alpha \leqslant 1$, then

$$H_1 = \frac{1-\exp(-\alpha pT)}{p} \text{ and } H_2 = \frac{1-\exp(-(1-\alpha)pT)}{p}$$

and if we put $a = 1/\tau_1$ and $b = 1/\tau_2$, then $F_1 = a/(p+a)$ and $F_2 = b(p+b)$

We can derive the following.

$$1-F_1H_2(z) = \frac{[z-\exp(-\alpha aT)]}{[z-\exp(-aT)]}$$

and

$$F_1H_1(z,\tau) = \frac{z[1-\exp(-\alpha aT)]}{[z-\exp(-aT)]}$$

and if $b \neq a$

$$F_1F_2H_1(z,\tau/2)$$

$$= 1+\frac{b[z\exp-\alpha aT/2-\exp-(1-\alpha/2)aT]}{(a-b)(z-\exp(-aT))} - \frac{a[z\exp(-\alpha bT/2-\exp(-(1-\alpha/2)bT]}{(a-b)(z-\exp(-bT))}$$

and

$$F_1F_2H_2(z,-T/2)$$

$$= \frac{b[\exp(-(1-\alpha/2)aT)-\exp(-\alpha aT/2)]}{(a-b)[z-\exp(-aT)]} - \frac{a[\exp-(1-\alpha/2)bT-\exp-\alpha bT/2]}{(a-b)[z-\exp(-bT)]}$$

and if $b = a$

$$F_1F_2H_1(z,\tau/2) = 1 + \frac{-z^2[1+\alpha aT/2]\exp(-\alpha aT/2) + z[\{1-(1-\alpha/2)T\}\exp(-(1+\alpha/2)aT) + }{(z-\exp(-aT))^2}$$
$$\overline{+\{1+(1-\alpha/2)aT\}\exp(-(1-a/2)aT)] + \{-1+\alpha aT/2\}\exp(-(2-\alpha/2)aT)}$$

and

$$F_1F_2H_2(z,-\tau/2) = \frac{z[-\{1+(1-\alpha/2)aT\}\exp(-(1-\alpha/2)aT) + (1+\alpha aT/2)\exp(-\alpha aT/2)] +}{[z-\exp(-aT)]^2}$$
$$\overline{+(1-\alpha aT/2)\exp(-(2-\alpha/2)aT) - \{1-(1-\alpha/2)aT\}\exp(-(1-\alpha/2)aT)}$$

so it is possible to derive conditions for $b \neq a$ and $b = a$, as below.

$$\frac{\Delta v_{in}}{\Delta v_{out}}(a-b)(z-\exp(-\alpha aT))(z-\exp(-bT)) = z^2\{(a-b)(1+L)+bL\exp(-\alpha aT/2)-aL\exp(-b\alpha T/2)\} +$$
$$+z\{(b-a)(1+L)(\exp(-\alpha aT)+\exp(-bT)) - bL\exp(-\alpha aT/2)(\exp(-bT)+1)+$$
$$+aL\exp(-\alpha bT/2)(\exp(-bT)\exp(-a\alpha T)\exp(b\alpha T)+1)\}+$$
$$+\exp(-bT)\exp(-a\alpha T)\{(a-b)(1+L)+bL\exp(-a\alpha T/2)-aL\exp(-b\alpha T/2)\}$$

$$\frac{\Delta v_{in}}{\Delta v_{out}}(z-\exp(-\alpha aT))(z-\exp(-aT)) = z^2\{1+L-L(1+\alpha aT/2)\exp(-\alpha aT/2)\} -$$
$$-z[(1+L)(\exp(-aT)-\exp(-\alpha aT)) - L\exp(-\alpha aT/2)\{(1-\alpha T/2)\exp(-aT)+(1+a\alpha T/2)\}]+$$
$$+\exp(-(1+\alpha)aT)\{(1+L)-L(1-a\alpha T/2)\exp a\alpha T/2)\}$$

These equations look rather daunting. However, when appropriate approximations and conditions are applied, useful design guidelines are obtained. These are summarized in Tables 2 and 3.

Manuscript first received by the Institution on 16th May 1980 and in final revised form on 25th February 1981
(*Paper No.* 2117/*CC*348)

The Author

David Mercy graduated in physics from Birmingham University in 1960, and joined the Colour Television Section of EMI Research Laboratories, where for the next nine years he worked on circuit design, optics and signal processing. In 1969 he transferred to the Broadcast Equipment Division and was involved in the development of a range of vision switching and mixing equipment. In 1974 he joined EMI's Cable Television group and was Development Manager of this group from 1975 until the transfer of the business to Jerrold International in 1979. Since that date Mr Mercy has been with the Radiation Department of EMI Electronics at Hayes.

UDC 621.391.823: 621.396.391.029.62
Indexing Terms: Signal interference, Radio systems, mobile

The magnitude of urban and suburban v.h.f. man-made radio noise

J. D. PARSONS, M.Sc., C.Eng., M.I.E.E., M.I.E.R.E.*

and

A. U. H. SHEIKH, Ph.D., C.Eng., M.I.E.E.†

SUMMARY

Measurements of the amplitude of the detected envelope of man-made noise have been made at a frequency near 80 MHz, and the results presented in the form of amplitude probability distribution curves. The major contributing sources are the ignition systems of vehicles and measurement locations such as business areas, arterial roads, and suburban areas have been chosen to give a variety of conditions. Although traffic density is the major parameter that has to be taken into account, the pattern of traffic flow is also a contributing factor.

* *Department of Electronic and Electrical Engineering, University of Birmingham, P.O. Box 363, Birmingham, B15 2TT*
† *Formerly at the University of Birmingham; now with the Faculty of Engineering, University of Garyounis, Benghazi, Libya.*

1 Introduction

The increasing demand for mobile radio channels for analogue and digital transmissions makes it important to provide a basis for predicting the performance of communication systems operating in built-up areas. There are several factors which have to be taken into account, for example, it is well known that the signal received at a vehicle moving in a built-up area is subject to severe fading, and this aspect of the problem has received much attention.[1] However, fading is not the only problem; mobile radio systems have to operate in areas which are heavily polluted with electrical noise and it is this aspect which will be treated here.

A mobile radio system is beset with noise from various sources and each source may have different characteristics. Firstly there is receiver noise which is Gaussian in nature and arises from the receiving system itself. Receiver noise is usually expressed in terms of nkT_0B, n being the factor by which the total receiver noise exceeds ambient noise. Atmospheric noise may be present, but it decreases rapidly with frequency and is generally negligible in the v.h.f. range. Galactic noise is also insignificant in the v.h.f. band as it is well below the background noise. By far the most important source of noise as far as mobile communication is concerned is that radiated by electrical equipment of various kinds. This noise, commonly termed 'man-made' noise is impulsive in nature, and therefore has characteristics quite different from Gaussian noise.

There are several potential sources of impulsive noise which could play a role in mobile communication systems. The radio is often installed in a vehicle which is itself a source of noise due to its own ignition and other electrical systems and the vehicle commonly operates in urban, suburban and industrial areas where it is close to other noisy vehicles. There are other extraneous sources of noise such as power lines and neon signs, industrial noise from heavy current switches, arc welders and the like, and noise from various items of domestic electrical equipment. These may or may not be significant contributors in any specific situation. In practice the level of man-made noise varies markedly with location and time,[2,3] so from a limited series of observations it is only possible to derive typical values and obtain some estimate of the variability.

In urban areas it is generally conceded that vehicle ignition noise is a major source of interference to v.h.f. mobile radio systems and this paper is concerned with the measurement of that particular kind of noise. In particular we are concerned here with the amplitude of the noise, and although it does not give the complete picture, information about the amplitude is clearly of fundamental importance in noise characterization.

2 Noise Characterization and Measurement

In a previous publication[4] the authors have discussed in detail the characterization of man-made noise and have described a practical noise-measuring receiver. Suffice it

0033–7722/81/110591+07 $1.50/0

to say here that although Gaussian noise can be uniquely described in respect of its significance in communication systems by its r.m.s. value, man-made noise consists in general of a succession of impulses which have random amplitude and random time spacing and characterization is not straightforward matter. The situation is further complicated by the fact that the noise is non-stationary in the statistical sense and this is hardly surprising when we consider that, at a roadside location, the precise conditions under which measurements are made vary from minute to minute. This makes it essential to record carefully the exact conditions under which each measurement has been made so that the measured parameters can be correlated with the environmental conditions under which the measurement was carried out.

Measurements of the amplitude of the impulsive noise are presented in the form of an amplitude probability distribution (a.p.d.). This takes the form of a graph, plotted on Rayleigh probability paper, which shows the overall percentage of time for which the noise envelope at the output of a receiver exceeds any particular value. The ordinate is usually expressed in dB relative to kT_0B. The a.p.d. gives the 'first-order' statistics of the measured noise since it contains no information about how the time is made up, i.e. the number of noise pulses which exceed any given value, and clearly therefore it will need to be supplemented by additional information in a more complete noise model. However, in itself it is an important and fundamental parameter which gives a large amount of information and can be used to predict the performance of data communication systems subjected to impulsive noise.[5]

The noise measuring equipment described in Ref. 4 is simple in concept. Briefly it consists of a commercially available radio receiver with a selectable i.f. bandwidth, from which three outputs can be taken for recording on an f.m. tape recorder, as shown in Fig. 1. The measured noise figure of the receiver is 4·4 dB. The receiver was slightly modified by inserting an attenuator in the i.f. section to overcome problems of limited dynamic range. By using a variable attenuator it is possible to effectively position a 'window' having a width equal to the unmodified dynamic range of the receiver so as to view any desired portion of the input signal. The effect of receiver bandwidth, dynamic range and impulse response on the measurements have been discussed in Ref. 4.

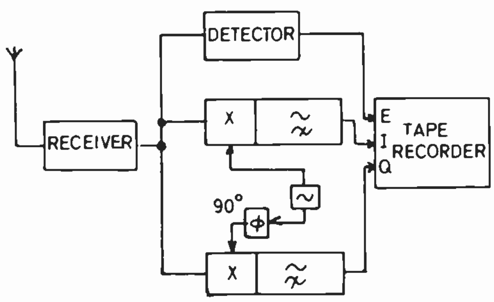


Fig. 1. Practical measurement system
E Envelope I, Q Quadrature synchronously demodulated outputs.

3 Field Trial and Data Analysis Procedures

Before conducting experiments, it was necessary to establish that any measured data were unlikely to be corrupted by coherent interference. For this purpose a spectrum surveillance was carried out, and several observations at various locations were made over periods which varied from 1 hour to over 24 hours. A spectrum analyser in the 'peak memory' mode was used for this purpose and it was found that there were no intelligent transmissions close to 80·1 MHz and this frequency was used for the measurements reported here. During the spectrum surveillance it was observed that there was a small percentage of very noisy vehicles which produced a noise level much higher than the average vehicle, a fact which has been reported previously.[6] These occasional bursts of high level noise are not apparent in the a.p.d. results presented in this paper but are of considerable importance in mobile radio reception.

For the field trials the experimental equipment was housed in a mobile laboratory which was driven to the chosen test site and parked alongside the kerb in the normal way. The vehicle engine was switched off and the receiving equipment was powered by an inbuilt 250 V, single-phase a.c. generator which had been suppressed so that it produced negligible r.f. noise. The receiving equipment was connected to a standard $\lambda/4$ whip aerial on the roof of the vehicle, the dimensions of the vehicle being such that the base of the aerial was 2·8 m from the ground. The receiver was tuned to 80·1 MHz with the aid of a digital counter which reads the local oscillator frequency and a digital automatic frequency control (d.a.f.c.) system was used to keep the receiver on tune. Before each test the receiver calibration was checked with and without the i.f. attenuator. For the tests reported here the receiver i.f. bandwidth was set to 20 kHz. Although it might appear desirable to use a much higher bandwidth, such as the standard 120 kHz used for quasi-peak measurements in this frequency range, this is not a realistic proposition. There is some difficulty in finding a suitable channel free from coherent interference and in addition the cost and time involved in processing wideband information is prohibitive. The use of an i.f. bandwidth of 20 kHz represents a compromise between the i.f. bandwidth used in the normal mobile systems which operate in this part of the spectrum, and the ideal.

The three outputs of the measuring system E, I and Q were connected to three channels of the f.m. tape recorder and two recordings of 6 minutes duration were made, one with an i.f. attenuator setting of 20 dB (upper window) and the other with zero attenuation (lower window). During each recording the number of vehicles passing the mobile laboratory was counted so that a check could be made on the correlation between the measured results and the environmental conditions. After completion of the trials the tapes were taken back to the laboratory for analysis, and the results reported here have all been obtained from an analysis of the envelope (E) channel alone.

The method of analysis to obtain the a.p.d. was to replay the recorded tape, and to sample and digitize the data using an a/d converter associated with a PDP-11 minicomputer. The a.p.d. can then be computed by counting the number of samples that exceed any preset threshold level. This was achieved directly without the necessity to store the individual digitized samples, by the use of software that incremented the count in an appropriate voltage bin within the computer memory each time a sample was taken.

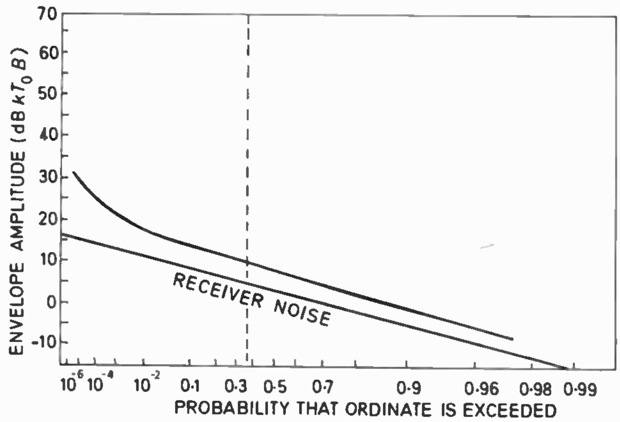


Fig. 2. Amplitude probability distribution for background noise.

4 Impulsive Noise Measurements

4.1 Choice of Locations

In order to assess the variability of the noise from one location to another within the City of Birmingham, the experimental sites were selected to provide a wide variety of conditions, e.g. business areas, major arterial roads, urban and suburban areas, and to obtain a spread of traffic densities. In built-up areas the traffic conditions vary with time, so at all locations several measurements were taken at different times. Suburban locations were also chosen on the basis of traffic density and the average density encountered during experiments was found to lie between 40 vehicles per minute and almost zero over the locations where the man-made noise was observed. At each experimental location, significant features such as the proximity of traffic lights, the type of street, whether dual carriageway or not were noted, as was the distance between the mobile laboratory and the centre of the road. The locations used are summarized in Table 1.

Table 1. Map references of measurement locations

Location	O.S. Map Reference
Bristol Road South	SP 018793
Broad Street	SP 060865
Colmore Row	SP 070871
Corporation Street	SP 071869
Great Charles Street	SP 065871
Hagley Road	SP 021859
Lordswood Road	SP 024853
M5	SP 965770
New Street	SP 069868
Stratford Road	SP 085851

The first step in the measurement programme was to check the level of background noise. A quiet spot at the rear of the University's Electrical Engineering Department was found and the background noise was recorded at 21.30 hours. An external a.c. power source was used instead of the mobile laboratory generator, so that all possible additional noise sources were eliminated. The noise data were analysed for a.p.d. and the result, shown in Fig. 2, forms a basis with which other results can be compared. The median noise level during this test was 11 dBkT_0B, which is almost identical with the value given for 'rural' areas in Ref. 7.

4.2 Measurements at Suburban Locations

The noise data typical of suburban areas is taken from measurements conducted at two locations, Lordswood Road and Bristol Road South. Lordswood Road is fairly wide, situated in a residential area, and the traffic density is normally fairly low. During the periods when experimental observations were made at this location the traffic density varied between 8 and 15 vehicles per minute with 11 being a typical value. Bristol Road South is also in a residential area, but it carries more traffic. The traffic density at this location varied between 12 and 20 vehicles per minute with 17 being typical. At both locations traffic flowed freely and was not influenced by road features such as roundabouts or traffic lights. The noise measured at these locations is presented in Fig. 3, which gives an impression of the difference between the

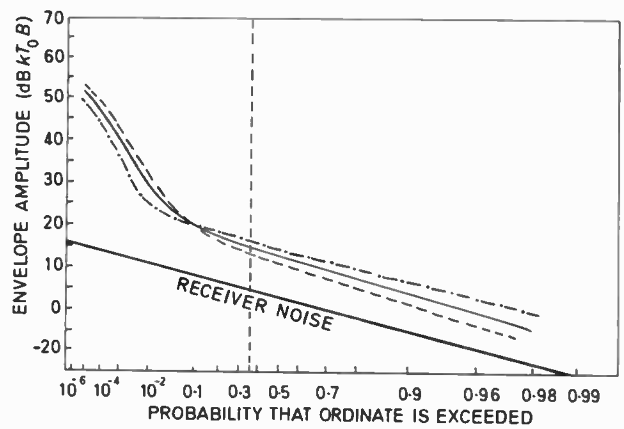


Fig. 3. Amplitude probability distribution for suburban areas
–·–·–·– Curve representing measurements in Bristol Road South (Average traffic density 12–20 vehicles per minute)
– – – – – Curve representing measurements in Lordswood Road (Average traffic density 8–15 vehicles per minute)
———— Generalized curve representing suburban locations (Typical traffic density 14 vehicles per minute)

two locations. The major influencing factor is, of course, the traffic density and it can be seen that the typical a.p.d. curve computed from measurements at various times on Bristol Road South, which generally carries more traffic, tends to have a higher median value and for much of the time lies above the curve derived from observations on Lordswood Road. The solid line in the centre represents a 'mean' and has been derived from all the measurements made at these two locations. It may be taken as typical of a suburban location where traffic flows freely at about 14 vehicles per minute, the traffic being mainly private motor vehicles and small vans with very few heavy lorries. Typically the median level of noise at these locations is 14 dBkT_0B with a 10^{-4} probability of exceeding 46 dBkT_0B.

4.3 Measurements at Urban Locations

Measurements were taken at several urban locations, the three presented here being on major arterial roads within a mile of the city centre but not within the central shopping area. The traffic is heavy at all locations, there being far more commercial traffic on these roads than in the suburban areas. Stratford Road, Broad Street and Hagley Road are all wide, single-carriageway roads carrying traffic to and from the city centre. The traffic density at all three locations is similar and varies between 25 and 34 vehicles per minute. Traffic lights are sited in Stratford Road and Broad Street, but the monitoring vehicle was positioned well away from them and the flow of traffic was continuous with few gaps and little evidence of any significant bunching.

Various experimental curves derived from measurements at these locations are shown in Fig. 4. There is a tendency for higher levels of noise to be measured when the traffic density increases but there are

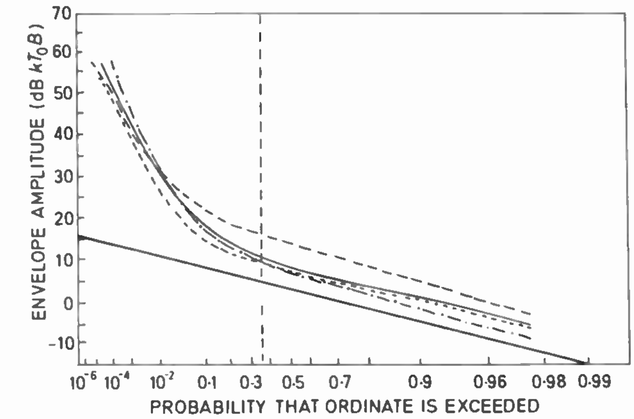


Fig. 4. Amplitude probability distribution for urban areas
- - - - - Curve representing measurements in Hagley Road (Average traffic density 30–34 vehicles per minute)
-·-·-·- Curve representing measurements in Stratford Road (Average traffic density 25–32 vehicles per minute)
········ Curve representing measurements in Broad Street (Average traffic density 29–33 vehicles per minute)
——— Generalized curve representing urban areas (Typical traffic density 30 vehicles per minute)

no significant differences between the a.p.d. curves for the three locations and clearly traffic density is the dominant feature influencing the measurements. The solid central curve may be taken as typical of noise measured in an urban area where traffic, consisting of private cars, vans and some heavy commercial vehicles flows fairly freely at an average rate of about 30 vehicles per minute. The typical median noise level at these locations is 9 dB kT_0B with a 10^{-4} probability of exceeding 52 dB kT_0B.

4.4 Measurements in the City Centre

Various locations in and around the city centre were selected on the basis of the expected traffic density and the type of street. The measurements presented here were taken on different dates and times in New Street, Corporation Street and Colmore Row, these being one-way streets located in the central business area of the city. They are identified on the map shown in Fig. 5.

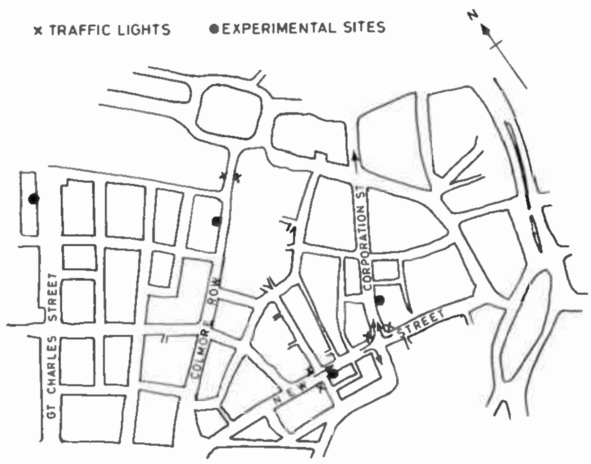
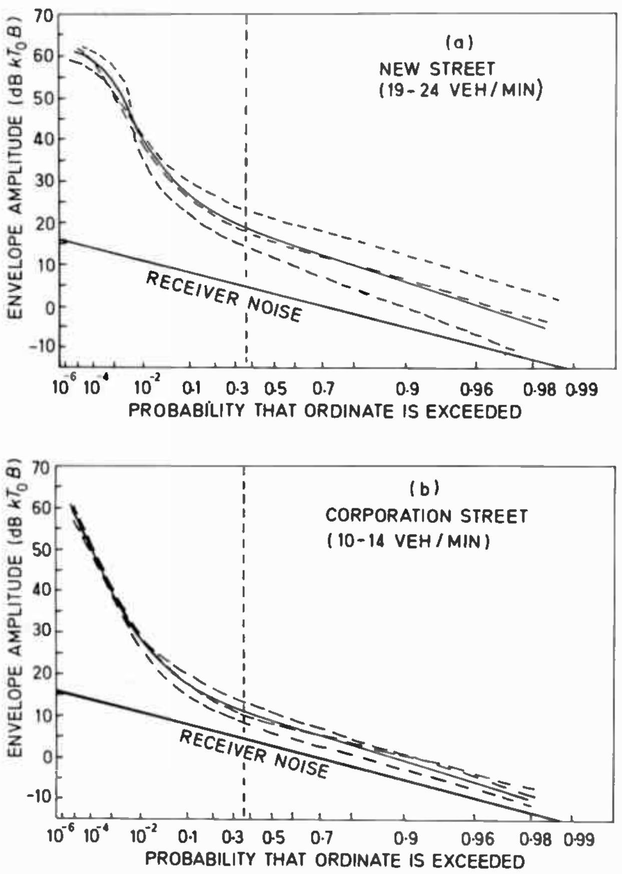


Fig. 5. Measurement locations in the City Centre

When measurements were made in New Street the mobile laboratory was parked in a meter bay just past a junction where the traffic was controlled by a set of traffic lights. Due to the proximity of these lights, the monitoring equipment was always influenced by the traffic. When the lights were green, vehicles were passing the mobile laboratory at a speed estimated as between 16 and 24 km/h (10 and 15 mile/h), thus contributing more noise than when the traffic lights were red and the vehicles were stationary and bunched about 10 m away from the mobile laboratory. The results at this location are summarized in Fig. 6(a). The experimental curves represent conditions encountered at different times of day and under different traffic conditions. Once again the solid line represents an average a.p.d. which may be taken as typical of this location. The average traffic density in New Street varied between 19 and 24 vehicles per minute and although there were many cars and vans, there were much fewer heavy commercial lorries in the



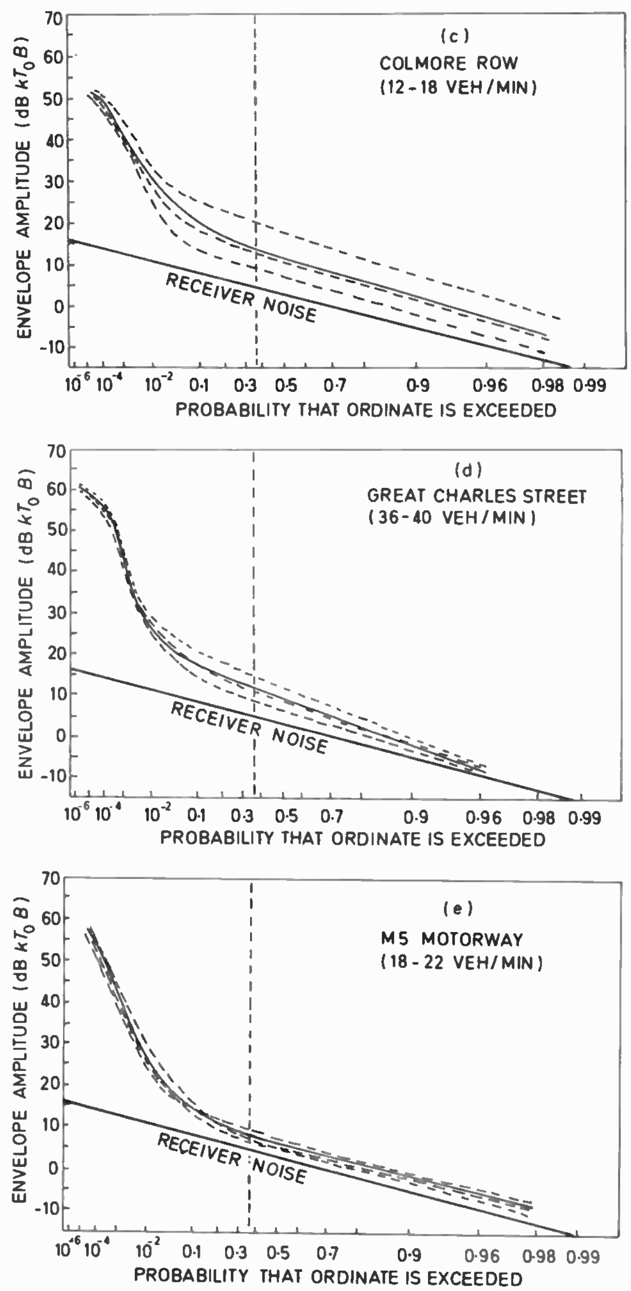


Fig. 6. Amplitude probability distribution for various locations
------ Experimental curves
——— Generalized curve representing average conditions at the locations indicated (d) and (e) represent average conditions for an urban expressway and an urban motorway respectively.

city centre than on the arterial roads previously described. The median noise level in New Street is about 17 dBkT_0B with a 10^{-4} exceedance probability at 59 dBkT_0B.

In Corporation Street the average traffic density is lower than in New Street, varying between 10 and 14 vehicles per minute, and the traffic pattern is also rather different. The entrance to Corporation Street is controlled by traffic lights approximately 100 m from the monitoring vehicle. The traffic therefore passes the mobile laboratory in bunches of 10 to 20 vehicles and there is then a relatively quiet period when the traffic lights are red and the vehicles are stationary some distance away. This, together with the generally lower traffic rate, produces a lower median noise level, 9 dB kT_0B in Corporation Street, compared with 17 dB kT_0B in New Street, as shown in Fig. 6(b). The noise level exceeded with 10^{-4} probability is also lower, at 51 dB kT_0B.

At the third central location, Colmore Row, the traffic density is intermediate between that of New Street and Corporation Street at 12 to 18 vehicles per minute. There are traffic lights along Colmore Row but the mobile laboratory was not positioned close to them and since there are several points at which traffic can enter and leave this particular street, there was a generally smoother flow of traffic and less evidence of severe bunching. The experimental curves and the generalized a.p.d. in Fig. 6(c) show a median noise level of 12 dB kT_0B midway between the levels measured in New Street and Corporation Street and at an exceedance probability of 10^{-4}, the level is 50 dB kT_0B, also an intermediate value. It therefore seems reasonable to suggest that the a.p.d. for Colmore Row should be taken as representative of business areas with the results for New Street and Corporation Street being indicative of the variations caused by different traffic conditions.

4.5 Measurements with Fast Moving Traffic

The measurements reported in the previous sections were all conducted at locations where the traffic was subjected to a 30 miles per hour (48 km/h) speed limit. It was therefore decided to make some measurements at locations where the traffic moved more quickly and two locations were selected, Great Charles Street, and the M5 motorway south of the city.

Great Charles Street is near the city centre, at the edge of the central business area. It is marked on the map in Fig. 5. Traffic is heavy and moves at speeds up to 80 km/h since this is part of the road system which feeds and links the motorways and is perhaps best described as an expressway. The traffic density in Great Charles Street was up to 40 vehicles per minute, the highest observed at any of the measurement locations, and the traffic flow was continuous with no gaps. The spread of a.p.d. graphs shown in Fig. 6(d) is noticeably narrower at this location, presumably due to the steady flow of traffic at all times of day. The median noise level is 7 dB kT_0B and the level exceeded with 10^{-4} probability is 52 dB kT_0B.

For measuring noise on a motorway, the monitoring vehicle was parked on the hard shoulder of the M5 and its position relative to the traffic was therefore similar to that for all the other measurements. Experimental a.p.d.s together with an average curve are shown in Fig. 6(e) from which it can be seen that once more there is no great variation with time, again probably due to the continuous flow of fast traffic at all times. The median noise level on the motorway is 10 dB kT_0B and the 10^{-4} probability level is crossed at 56 dB kT_0B. The traffic density on the motorway is similar to that along Bristol Road South, but the shape of the a.p.d. curve is quite different, the curve for the M5 having a lower background level, but rising very steeply between the envelope amplitude levels of 25 and 50 dB kT_0B. The difference is undoubtedly due to the different vehicle speeds and vehicle spacings involved. On Bristol Road South vehicles are travelling at about 48 km/h and are spaced at a distance appropriate to that speed. Because of the relatively low speed there are always a few vehicles fairly close to the monitoring vehicle so the background level is high and the difference between the median and 10^{-4} probability values is not very large, 31 dB. On the M5, on the other hand, vehicles are moving at 110 km/h (70 mile/h) (so producing more noise) and are spaced much further apart. Because of this high speed the vehicles are only within range of the monitoring vehicle for a short time and this results in a low level of background noise with short bursts of high level noise. This gives the very steep slope at low probability levels, characteristic of this kind of traffic pattern. The difference between the median and 10^{-4} probability levels is 46 dB.

5 Discussion

The results of the various observations made at the different locations are summarized in Table 2. An examination of this Table together with the a.p.d. curves reveals significant differences between the results obtained.

Table 2. Summary of results

Location	Designation	Average Traffic Density (veh/min)	Level exceeding given probability (dB kT_0B) $p=0{\cdot}5$	$p=10^{-4}$
Lordswood Road Bristol Rd. South	suburban	14	14	46
Stratford Road Broad Street Hagley Road	urban	30	9	52
New Street	business	19–24	17	59
Corporation St.		10–14	9	51
Colmore Row		12–18	12	50
Gt. Charles St.	urban expressway	36–40	7	52
M5	urban motorway	18–22	10	56

The suburban results in Fig. 3 show that as far as background noise is concerned Lordswood Road is marginally noisier than Bristol Road South. However the typical curves for these locations cross at about 0·1 probability and those parts of the curves representing impulsive noise are nearly parallel with Lordswood Road being a few dB lower.

As far as urban measurements are concerned there are some striking similarities between the shapes of the a.p.d. curves and the measured probabilities at various envelope levels for Broad Street and Stratford Road which carry similar traffic flowing in an almost continuous stream. The background noise on Hagley Road is higher, but at low probabilities all three curves are very similar. The generalized a.p.d. curve for urban areas reaches higher noise levels than the suburban curve and the slope in the low probability regions is also higher.

In the centre of the city we encounter a different situation in which vehicles move in bunches under the control of traffic lights. In many cases, e.g., New Street, Fig. 6(a), the a.p.d. curve is very steep at low probability levels and we see the influence of another factor, traffic pattern, at these locations.

If we examine the results for the heavily built-up city centre area we see significant differences in the a.p.d. curves measured, for example, in New Street and Corporation Street. From a comparison of Figs. 6(a) and (b), it is apparent that the probability distribution curve has been displaced towards the left in the case of Corporation Street although the shapes of the curves are not very different. In the case of New Street, the background noise is higher by 8 dB for a probability of 0·5 and New Street is noisier than Corporation Street by approximately 12 dB and 6 dB at probability levels of 10^{-3} and 10^{-4} respectively. The reason for the increased background noise in New Street is most probably the noise contributions from the increased numbers of vehicles, because the average traffic rate is 60% more than that in Corporation Street. The slope of the a.p.d. curve between 10^{-2} and 10^{-4} probabilities is greater in the case of Corporation Street, which suggests a greater variability in the level of noise. This may be due to the rather intermittent arrival of vehicle groups, as the entrance to Corporation Street is controlled by traffic lights, whereas the traffic flow in New Street, although also controlled by lights, is smoother.

Traffic pattern also has an influence as far as the expressways and motorways are concerned, as can be seen from Figs. 6(d) and (e). The general shape of the a.p.d. curves is very similar particularly in the low probability region. The traffic density on the M5 is similar to that in New Street, but the pattern of traffic flow is clearly very different, and the 'break-point' on the a.p.d. curve for the M5 is an order of magnitude lower on the probability axis. It seems clear that both traffic density and traffic pattern influence the shape of the a.p.d. curve. Type of traffic is also likely to have an effect.

The designation used above to describe environments follows as far as possible that of CCIR,[7] where measurements of man-made noise in the United States were used to produce median values for various locations as a function of frequency. It is worth noting that the measurements reported in this paper have produced median noise levels consistently lower than the CCIR values in business and residential areas, the difference being of the order of 10 dB as suggested previously.[4]

6 Conclusions

An extensive series of field trials, of which those reported here form a part, have been used to derive information about the envelope amplitude of v.h.f. impulsive noise in various environments. The results have been presented in the form of the a.p.d., a useful first-order statistic which gives the probability that the amplitude of the detected noise envelope exceeds any given level.

Examination of the a.p.d. curves measured in different circumstances show that they all have the same general shape, a region of low slope at lower amplitude levels and a region of high slope at higher amplitude levels. If we were to produce an approximation to the a.p.d. by two straight lines the break-point between the two regions would occur in the probability range 10^{-3} to 10^{-1}. Heavier traffic conditions cause the break-point to move to the right, i.e. to occur at a higher probability level. In suburban areas where the flow of traffic is relatively smooth the transition between the two regions is gradual. However, in the city centre, situations often exist in which vehicles move in bunches under the control of traffic lights and we then often see an a.p.d. curve which is very steep at low probability levels with a much sharper transition between the two regions. The conclusion therefore is that although traffic density is a major factor in determining the a.p.d., the detailed shape of the curve obtained at any specific location is a function of factors such as the speed of vehicles and the proximity of traffic lights which have an influence over the pattern of traffic flow.

Finally, although, by the nature of the work, experimental results were never precisely repeatable, returning to the same location on other days and at other times usually produced results, under similar traffic conditions, which were sufficiently similar to previous measurements to justify the claim that the results presented here are typical of the situation they represent.

7 Acknowledgment

This work was based on studies carried out for the UK Home Office, Directorate of Telecommunications (their report TM8) and the authors wish to thank the Director for permission to publish.

8 References

1 Jakes, W. C. (Ed.), 'Microwave Mobile Communications' (Wiley, New York, 1974).

2 Herman, J. R., 'Survey of man-made noise', *Progress in Radio Science*, **1**, pp. 315–48, 1971.

3 Horner, F., 'Techniques used for the measurement of atmospheric and man-made noise', *Progress in Radio Science*, **2**, pp. 177–82, 1971.

4 Parsons, J. D. and Sheikh, A. U. H., 'The characterization of impulsive noise and considerations for a noise-measuring receiver', *The Radio and Electronic Engineer*, **49**, no. 9, pp. 467–76, 1979.

5 'World Distribution and Characteristics of Atmospheric Radio Noise.' CCIR Report 322–1.

6 Shepherd, R. A., Gaddie, J. C. and Shohara, A., 'Measurement Parameters for Automobile Ignition Noise', Stanford Research Institute, Final Report SRI 7502-C-2-10, June 1975.

7 'Man-made Radio Noise', CCIR Report 258-3.

Manuscript first received by the Institution on 26th November 1980 and in revised form on 8th April 1981.
(Paper No.2118/Comm 328)

Symposium Report

Telemetric Studies of Vertebrates

A Report by R. B. Mitson and C. L. Cheeseman (Ministry of Agriculture, Fisheries and Food) on a Symposium organized by the Mammal Society with the co-operation of the IERE Aerospace, Maritime and Military Systems Group

With his inferior agility and senses, man has, until recently, been unable to come to terms with small, shy, or secretive animals. A meeting at the Zoological Society of London in November 1980 showed that significant advances in knowledge of the whereabouts, activity and in some cases the physiology of animals are being made. These are due to the use of radio or underwater acoustic telemetry techniques as appropriate when applied to bat, elephant, cod-fish and many others. The purpose of this report is to outline some of the problems in meeting the needs of investigators into animal behaviour and to report briefly on a few of the successful applications.

Engineering aspects of telemetry occupied the first session of the two-day symposium and the remaining three covered a variety of applications to fish, birds and mammals of many species. It was evident that biologists usually identified areas of research likely to benefit from the use of telemetry, then asked an engineering science group to recommend or design a system. However, in some instances, because of limited financial resources they tackled the construction of transmitter tags themselves, but bought receivers and aerials.

Radio tags are of the 'pinger' type, typically giving a pulse transmission of 80 ms duration at a rate of one every two seconds, but a special exception are the satellite interrogated transponders attached to far-ranging animals such as polar bears.

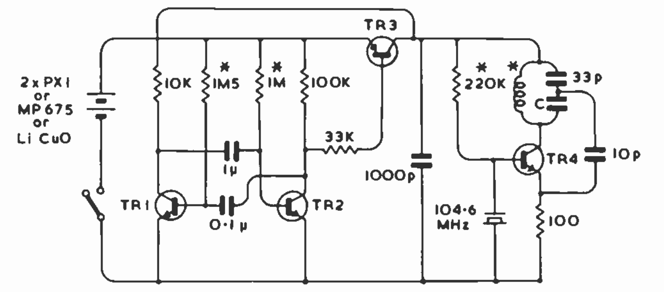
Underwater acoustic tags are often 'pingers' having a pulse duration of about 3 ms, transmitted at two second intervals. Transponders are also used where the precise range from ship to fish is an advantage and the signal is to be displayed in relation to sea bottom features. Where physiological or other variables are telemetered from terrestrial or aquatic animals, either pulse rate, width, or position modulation is used.

Before the advent of the transistor it was impractical to attach electronic units (tags) to any except the largest free-ranging animals. Even when transistor technology was well established, the size of simple transmitters restricted their use to fully-grown adults of most species, because the other necessary components had not then been similarly miniaturized. Electronics technology in itself now poses no problems, the size and weight of complex circuits for animal telemetry are insignificant compared to batteries, aerials and transducers and in fact very few tags benefit from the latest technology. The important factor is numbers; animal tags cannot easily be standardized and even if they could, the numbers required, in terms that the electronics industry understands, would still be minute. Costs therefore are prohibitive for advanced methods of construction so a compromise is made based on a number of factors, for example weight and size relative to the animal, length of operating life, and for radio, the need to conform to the regulations governing transmissions.

Despite difficulties, progress is being made and the first paper at the symposium was a review of work in North America with the accent on engineering aspects. A small number of laboratories in the USA are dealing with telemetry for species as diverse as the polar bear and the alligator (at the large end of the scale), but animals as small as mice are also studied using radio. Much of the equipment comes from commercial sources, but this is a 'cottage' industry comprising small firms with few engineering staff and little money for development.

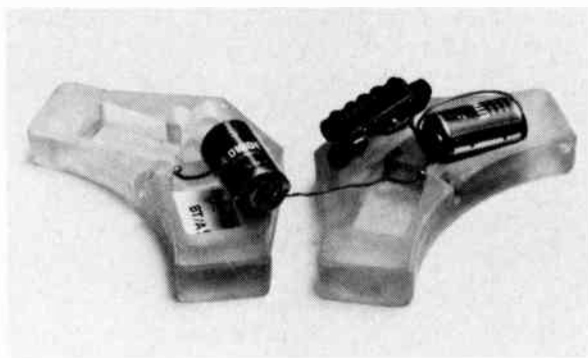
Frequency of Operation

In the USA a wide range of frequency bands is used for radio tracking of wildlife, although those authorized by the Federal Communications Commission are limited to 40·66–40·7 MHz and 216–220 MHz. Many groups who use other bands have 'special use' permits. In the UK, Home Office regulations limit the bands to 104·6–105 MHz and 173 MHz, but the latter is not a protected band. A paper at the meeting dealt with the UK regulations, pointing out that provided the allocated frequencies were used, there was very little possibility of interference and if the equipment met Specification MPT 1312 spurious responses would not occur in adjacent bands, nor would the tags drift off



TR1, TR2 BCW33R
TR3 BCX 18
TR4 BFS 20R
C 12p (fish tag) or 8·2p (animal tag)
* Select as required

(a)



(b)

Fig. 1. (a) Diagram of the basic radio transmitter circuit used by many workers in the UK (104 MHz). It was due to the late G. E. Ashwell of the Worplesdon Laboratory, Ministry of Agriculture, Fisheries and Food. (b) An exploded view of the radio tag circuit board, aerial, battery and case, weighing 95 g, for use by MAFF on badgers.

frequency (a problem complained about by users of some imported equipment).

The main disadvantage with the lower frequencies is due to the size of the receiving aerial, which many of the investigators must carry, whatever the terrain. Imagine struggling through dense woodland with a three or more element Yagi in one hand whilst trying to operate a receiver with the other! For 104 MHz many groups in the UK use collapsible H-Adcock aerials which can be set up on a sturdy tripod and will give a null indication of direction to an accuracy of about ±5°.

Transmitter Power

Little information was given about radiated power from transmitters because few people possessed the equipment to measure it. One UK paper included details of performance checks to ensure that tags complied with MPT 1312 for level of radiated power and spectral distribution, frequency stability, pulse duration and repetition rate. Although a common circuit design is used, Fig. 1 (a), the power output varies according to the battery pack selected for a particular animal study. When run from a low voltage the power output is typically about 40nW, but even with a larger battery it does not exceed 150nW. This design, Fig. 1 (b), has a small iron-dust rod aerial wired directly to the circuit board, the whole assembly being enclosed within a case cast from epoxy resin whose dimensions are determined by the physical characteristics of the particular animal to be studied. Badgers, foxes and coypu have been fitted with these transmitters. It seems unlikely that any radio tags will approach the permitted power level of 1 mW of this class of transmission, because the aerials are less than 1% efficient and it is necessary to conserve battery power to allow the studies to continue over periods of 9 months or a year. It is interesting to compare power requirements for radio tags and underwater acoustic tags. Despite transducer efficiencies of 80% for the latter, about 1W input is needed to achieve a range of 500 m at 300 kHz, whereas 1 mW of electromagnetic energy at v.h.f., radiated in air, can be detected some tens of kilometres distant.

Encapsulation

In protecting the tag from the environment it is necessary to ensure that an absolute minimum of weight is added, but no ingress of moisture is tolerable. Problems of case strength or encapsulation arise when animals practise mutual grooming and possibly try to remove transmitters by determined biting; it is fortunate that the strength of bite varies directly with size of animal. Direct encapsulation of radio circuits has the effect of changing the tuning and may also introduce mechanical stress to the construction.

Applications

Both marine and freshwater fisheries research was reported, in the former case acoustic transponders were used and in the latter, radio 'pingers'. Recent work has enabled the heart-rate of free-swimming plaice in the sea to be measured in real-time by a modified transponding system used in conjunction with an electronic sector-scanning sonar. The tag used for this purpose was simple compared to one which carried a miniature magnetic compass so that the orientation of the fish could be determined at any time, Fig. 2 (a) and (b). Pulse-position modulation was used to convey the information on direction. These studies are concerned with the energy expended during swimming and the mechanism of navigation on migration, respectively.

Radio tags have several advantages over acoustic tags for tracking fish in freshwater, especially in aerated and turbulent conditions. Work was described on the movements of salmon in rivers using a 104·6 MHz tag in the stomachs of these fish which could be detected and followed at ranges up to 1·5 km. The purpose was to monitor the reactions of migratory fish to natural and regulated fluctuations in river flow.

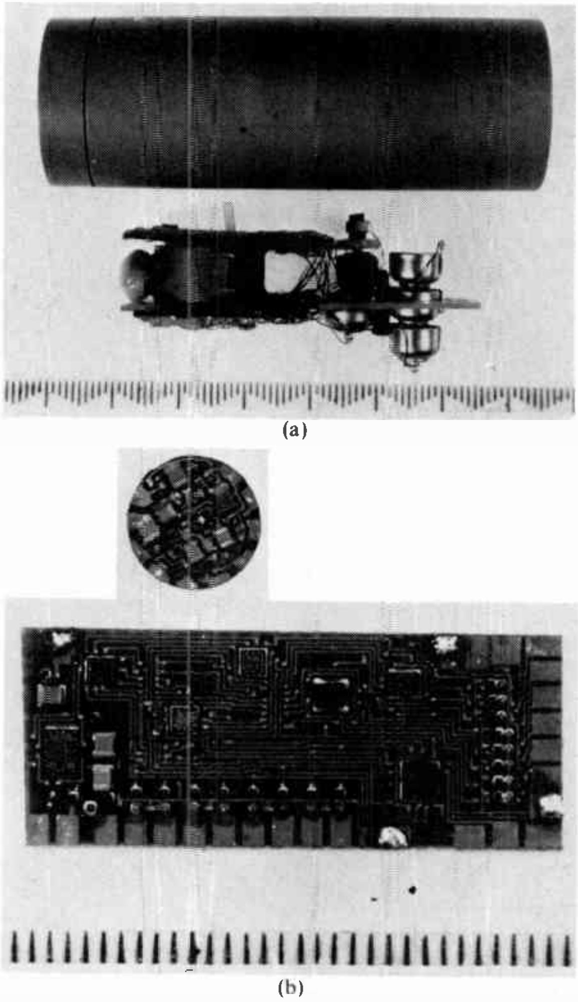


Fig. 2. (a) A fish compass tag case shown with the component assembly below. The complete unit weighs 8 g in water. Scale in mm. (b) Circuit board showing the hybrid construction used for the compass tag: top disc is the transponder; lower board is compass circuit logic. Scale in mm. (Courtesy of Ministry of Agriculture, Fisheries and Food.)

In both radio and acoustic telemetry the specification of the telemetry device is determined by a number of factors including the animal's size, anatomy, method of locomotion and general behaviour. In detailed studies of animals using telemetry there would be little point in doing anything which might affect their subsequent behaviour. Most authors seemed acutely aware of the problem and took great care to fit transmitters properly, with the least discomfort to the animal. In those cases where it was possible to compare the results obtained from control animals, no serious observable differences were reported. Speakers described a number of methods of attachment, the most common for terrestrial mammals being the use of a collar. In some cases the collar even formed the transmitting aerial.

A tag weight of not more than 5% of the body weight is aimed for, but of course, if adhered to in the larger species this would be unnecessarily bulky. Generally the principle seems to be, the smaller the better. With the smallest tags it is battery size, not electronic components which is the limiting factor.

With hedgehogs an encapsulated transmitter was glued to the

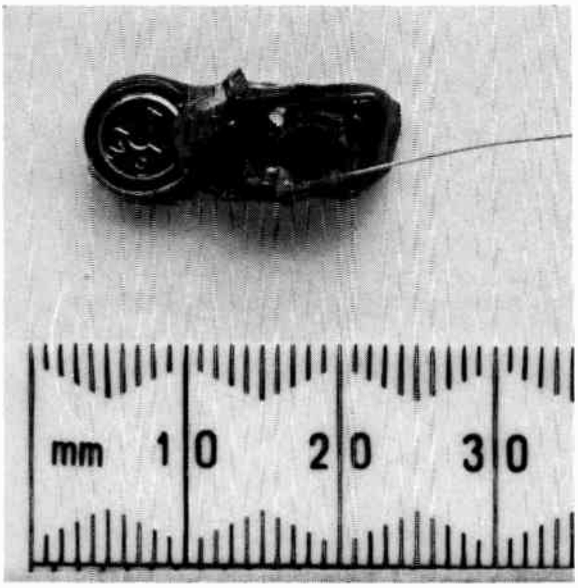


Fig. 3. Latest transmitter tag used for tracking Greater Horseshoe Bats. Its weight is now reduced to 1·2 g by the use of a zinc/air cell. The total aerial length is about 120 mm. (Photograph by courtesy of H. R. Arnold, Institute of Terrestrial Ecology.)

spines with epoxy resin. Although this method was not permanent (the spines gradually moult) there was no difficulty in periodically recapturing individuals under study to reglue the tag. The same method of attachment was used in a study of one of Britain's rarest flying mammals, the Greater Horseshoe bat. A transmitter of 1·5 g weight with a trailing wire aerial (Fig. 3) was glued mid-dorsally to the fur. The tag dropped off in a matter of days but as battery life was similarly short this was not a disadvantage. In the study many new roosts were located in sites previously considered unsuitable, and much new information was obtained on the preferred feeding habitats. For such work a tag weight of 1 g is desirable having a life of 10 days with a range perhaps 50% greater than the present 500–1500 m.

For larger animals, the size of the tag and associated aspects of performance such as range and battery life, are not so limiting. A study of leopards by radio telemetry illuminated the relationship of this large cat with other predators and prey species, giving an understanding of the niche occupied by the leopard in the Serengeti ecosystem. Using commercial radio tags mounted on collars made of machine belting, the sightings of individuals increased approximately thirty-fold. It was still necessary, however, to keep a discreet distance to avoid either disturbing the leopard or thwarting its hunting efforts.

A study of badgers in urban Bristol highlighted another of the problems which biologists face using telemetry, namely that of error in bearing accuracy caused by reflections of the radio signal at a particular site. Although badgers were rarely seen, close contact was maintained from nearby roads and paths (access into gardens obviously being restricted). Apart from the occasional noise of a badger turning over a dustbin, the radio tag provided the only means of detection.

An ingenious device used in the study of woodcock enabled the investigator to tell whether or not the bird was flying. A thermistor in the electronic circuit of the tag was placed under the wing of the bird. This cooled rapidly when the bird flew, and modulated the pulse interval of the transmitted signal. In this way it was possible to discover more about the strange 'roding' display of breeding male woodcock. Tags were attached mid-dorsally to the birds by means of an elastic harness which proved satisfactory except in the case of one bird whose beak become stuck under the harness during preening. A different method of attachment to birds was used in a study of goshawks, the tag being mounted on the tail feathers which were lost at the next moult. This method of tag detachment is desirable when batteries have expired and the animal cannot be recaptured easily.

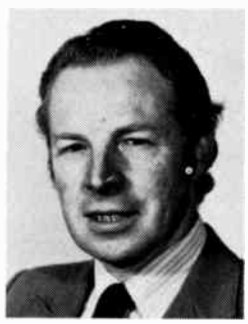
One paper described the only true example of a radio biotelemetry study presented at the meeting. An implanted tag transmitted an electrocardiograph signal from free-swimming diving ducks. This showed that the heart rate increased just before diving and again just before surfacing. There was no slowing of the heart rate during the dive contrary to the previously accepted principles of the physiology of diving ducks. It was also interesting to discover that the heart-beat rate was directly proportional to the leg-kick rate whilst swimming.

During the course of the meeting it was evident that there is considerable potential for the development of biotelemetry. It was also clear that successful application of telemetry demands skill and ingenuity from the operator in tackling the various operational problems and interpreting data. Telemetry has opened new horizons to biologists but there is a limit to the amount of information which can be automatically monitored. Ultimately there is no substitute for a pair of observant eyes, which at night can be aided by a beta light fixed to animals.

The Zoological Society of London are shortly to publish the Symposium papers. Details are available from the Assistant Director of Science, Zoological Society of London, Regent's Park, London NW1 4RY.

Contributors to this issue

Asrar Ul-Haq Sheikh graduated from the West Pakistan University of Engineering and Technology and subsequently obtained the degrees of M.Sc. and Ph.D. from the University of Birmingham in 1966 and 1969 respectively. He held lectureships at Universities in Pakistan, Iran and Libya between 1969 and 1975. In 1975 he returned to the University of Birmingham as a Research Fellow to work on the characterization of man-made noise. After the completion of this work in 1978 he returned to Libya and is now at the University of Garyounis, Benghazi.



David Parsons (Member 1967) graduated at University College of South Wales, Cardiff, in 1969 and later received the M.Sc.(Eng.) degree at King's College, London. He spent three years at the GEC Applied Electronics Laboratories, Stanmore and subsequently held teaching appointments at the Polytechnic, Regent Street, London, and Birmingham Polytechnic; he is now a Senior Lecturer in electronic and electrical engineering at the University of Birmingham. His current research interests in mobile communications cover propagation, noise and systems aspects. Mr Parsons is a member of the Committee of the West Midland Section of the Institution and has in the past held office as Chairman and as Secretary.

Colloquium Report

Waveguides and Components for the 80–300 GHz Frequency Range

A one-day Colloquium was organized by the Institution's Components and Circuits Group and held at Imperial College, London in April 1981. This report is by Professor D. J. Harris, Head of the Department of Physics, Electronics and Electrical Engineering at the University of Wales Institute of Science and Technology, Cardiff, who arranged the programme.

The upsurge of interest in the short-millimetric wavelength part of the electromagnetic spectrum was shown by the presence of about 80 participants at this colloquium. Under the joint chairmanship of Dr K. W. Gray of RSRE, Malvern, and Professor P. N. Robson of the University of Sheffield, ten presentations were given covering various aspects of propagation systems, and corresponding active and passive devices.

Work over many years at EMI Electronics, Wells, was described by M. W. Booton, M. C. Carter and E. G. Stevens. Dominant-mode rectangular guide systems and components using electroforming techniques are made and operated up to 300 GHz. Directional couplers for a 280 GHz system are constructed using etched copper foil between mandrels before electroforming. Branch guide 3 dB couplers, and 10 or 20 dB slot couplers with 30 dB isolation can be made. Rotary vane attenuators with an accuracy of 1 dB in 50, dielectric vane phase shifters, 5-pin band-bass filters with, for example, a 279–280 GHz passband and 5 dB insertion loss, and calorimeters for the 5–500 mW range were quoted. Single-ended and balanced mixers using GaAs Schottky diodes in Sharpless wafers were described, with conversion losses of 17 dB at 325 GHz and 8 dB at 80 GHz. Harmonic mixers, up-convertors, and switch-controlled phase and amplitude modulators have also been constructed. Electroforming techniques are also used for horn, dish and slot-array aerials. Quasi-optic techniques have been used for the 80 GHz–1 THz range. The need for compact lightweight alternatives has led to an investigation of planar technology guides. A comparison of different approaches was given and a 85 GHz radiometer in Duroid-based fin-line was described.

The advantages of E-plane circuitry, using various fin-line and strip-line configurations, was also stressed by S. J. Nightingale, R. N. Bates and M. D. Coleman of Philips Research Laboratories, Redhill. In all cases thin low-permittivity substrates are mounted between the broad walls of rectangular guide in the E-field plane, with conductor patterns on one or both sides of the substrate to form the transmission lines and circuit elements. Duroid ($\varepsilon_r = 2{\cdot}22$) is the main copper-clad substrate used. Complex conductor patterns are produced by photo-lithography. Unilateral fin-line losses between 0·02 and 0·2 dB per wavelength are obtained, depending on slot-width and frequency between 33 and 115 GHz. E-plane microstrip losses are greater, at about 0·7 dB per wavelength up to 75 GHz. A wide range of E-plane components has been made up to 90 GHz, including balanced mixers, Doppler mixers, p-i-n switches, directional couplers, attenuators and fin-line to waveguide transitions. Performance figures quoted were 2 dB insertion loss for p-i-n diode switches at 90 GHz, p-i-n attenuators with a 1–20 dB range, detectors up to 140 GHz with typical sensitivities of 50 mV per mW, and a balanced mixer for 60–90 GHz with 8·5 dB conversion loss. A metallized ABS plastic technique can give a lightweight product.

Contributions by M. J. Sisson, P. M. Wood and D. G. Monk of GEC Hirst Research Centre, Wembley, and by T. H. Oxley of Marconi Electronic Devices, Lincoln, both dealt with the extension of microstrip techniques to 100 GHz. The dielectric used is z-cut quartz of 100 μm thickness, with guide losses of 0·1–0·2 dB per mm at 140 GHz. A range of GaAs beam-lead mixer diodes has been developed with very long capacitance (e.g. 0·03 pF), low series resistance (e.g. 6 Ω) and good mechanical strength by providing a low-loss glass support wall around the basic chip. An estimated f_{co} of 2500 GHz was quoted for the DC 1346 diode, other parameters being $C_T = 0{\cdot}03$ pF, $C_j = 0{\cdot}01$ pF. A versatile micro-l.i.d. chip carrier has been developed for such applications as Si detectors, low barrier height GaAs diodes, high burn-out devices and noise diodes. Beam lead diodes are about 800 μm in length, whilst the micro-l.i.d. chip carrier is about 740 μm in length and 500 μm in height. Circuit elements such as stepped-ridge transitions, couplers and r.f. filters have been developed, and the range of components includes single-ended and balanced mixers, harmonic mixers, detectors, s.s.b. up- and down-convertors and noise sources. A rectification efficiency of 0·6 mA per mW at 140 GHz for a single-ended mixer was reported, and a conversion loss less than 7 dB for a balanced mixer at 92–96 GHz with 6–10 mW l.o. power. It was contended that strip-line techniques now well established for the 20–40 GHz range could mostly be extended for application up to 100 GHz.

P. M. Brigginshaw, J. E. Curran and C. Brown of GEC Hirst Research Centre, Wembley, discussed the potential of ferrite device circulators in microstrip circuits for the 60–95 GHz range. Crystal quartz, z-cut, is used as substrate with a cylindrical or triangular section ferrite insert. The maximum saturation magnetization available is 0·5T rather than the optimum 2T. This limits bandwidth, e.g. to 4·5% for 20 dB isolation/return loss for a disk junction at 94 GHz. Both apex and side-wall coupled triangular junctions have been investigated. Apex coupling gives a broader bandwidth circulator, with typically 15 dB isolation/return loss over about 30% bandwidth at 80 GHz. Corresponding figures for side-wall coupling are 20 dB over 10%. Cylindrical inserts performed as well as triangular in general. For good overall performance well-matched, low-loss, broadband, stepped-ridge transitions for 60–90 and 75–110 GHz have had to be developed, as well as good terminations, obtained by deposition of nichrome directly on the substrate.

Two presentations dealt with alternative guiding structures, R. G. Gelsthorpe and N. Williams describing work at ERA Technology, Leatherhead, on dielectric guide and components. Two configurations, image and insular guide, were especially considered. The former consists of rectangular dielectric strip ($\varepsilon_r = 10$) mounted directly on a conducting ground plane, whilst in the latter the dielectric strip is separated from the ground plane by a second lower-permittivity ($\varepsilon_r = 2{\cdot}5$) dielectric strip. The main advantages of the latter are that conduction losses are reduced and problems at the dielectric/conductor interface are largely removed, but this is at the cost of increased complexity and an earlier onset of higher-order modes. The guide wavelength is only 10–40% reduced on the free-space value, and small radius bends are therefore prone to loss by radiation. A wide range of devices has been fabricated, including couplers, filters, isolators and antenna arrays, and their use has been demonstrated by the construction of a total transceiver circuit. Since dielectric image guide is planar it is amenable to construction by repeated thick film printing techniques. The requisite line geometry can be printed and fired, and work is under way to formulate a printing paste with more appropriate dielectric properties. Work on groove-guide at UWIST, Cardiff, was described by D. J. Harris, Y. M. Choi and S. Mak. Parallel conducting planes, with a groove in each in the direction of propagation, give a simple configuration with a loss about an order of magnitude less than corresponding rectangular guide, and which is large compared with the wavelength and yet can be single mode. No dimensions are critical to a fraction of a wavelength. Transverse dimensions are approximately 20λ by 3λ, and typical experimental losses are $0{\cdot}2\ \mathrm{dB\ m^{-1}}$ at 100 GHz. Coupling together of sections of guide is straightforward, without high precision, and the guide is therefore relatively easy to make and use. The guide operating bandwidth can be very large—e.g. an octave or more, and the upper frequency is expected to be in excess of 300 GHz. Bends, transitions to rectangular guide, couplers, and groove-guide detectors were described. The latter use GEC beam-lead diodes in triangular dipoles formed from Duroid and mounted across the guide with a cylindrical back reflector. Power handling of groove-guide will be very high compared with other guides at these frequencies.

Three contributions were concerned with millimetre wave sources, two on solid-state sources and one on millimetre wave magnetrons. Transferred-electron oscillators were discussed by I. G. Eddison, I. Davies and D. M. Brookbanks of Plessey Research (Caswell). Both InP and GaAs oscillators have been constructed. InP is expected theoretically to have a better power and efficiency capability at a higher frequency than GaAs, and this is confirmed experimentally. Fundamental frequency operation of GaAs t.e.o. is limited to about 60 GHz, whilst 30 mW is obtainable at 90 GHz with InP, and milliwatt power levels at 140 GHz. The oscillation spectrum is good. Harmonic mode operation using InP is expected to achieve 200 GHz, and fundamental mode operation at 100 GHz should give a power approaching 100 mW. Impatt oscillators were described by I. Davies, H. S. Gokor and D. M. Brookbanks of Plessey, Caswell. Silicon is still the most suitable material for millimetre wave impatts for high power and high efficiency, using material grown by vapour phase epitaxy. Both double-drift and single-drift devices can be used, but the former appear to have advantages in power and efficiency up to 100 GHz. Typical performance figures are 400 mW at 12% efficiency at 90 GHz using double-drift devices and 100 mW at 7·5% efficiency at 130 GHz with single-drift devices. The power limitation is largely set by thermal dissipation problems, and improved performance is obtained through the use of diamond heat sinks. Powers of 1 watt cw and 12 watt peak for pulsed devices at 90 GHz are anticipated. The design features of a 95 GHz production magnetron were described by Miss D. C. Platts of English Electric Valve Company, Chelmsford. The tube has a rising-sun anode, hobbed from a copper block, and a barium aluminate cathode. A peak power of 3 kW at 0·0001 duty cycle is available, with beam voltage and current of 10 kW at 7A. The use of samarium cobalt focusing magnets can reduce the overall weight to 1·8 kg. Also available is an 80 GHz magnetron of 5 kW peak power with frequency agility by mechanical tuning through a piezoelectric drive, giving 600 MHz frequency shift. The magnetrons are rugged and long-life for systems operation.

It was clear from the presentation and discussion that a significant amount of work, and considerable progress, is being made in the UK on millimetre wave components and systems up to 100 GHz. Sources, active devices, passive components and guiding systems are available using rectangular guide, strip-line, and E-plane circuits—especially in the fin-line form. Dielectric guide could find application for cheap mass-produced systems. Solid-state sources at these frequencies are made in the UK, although mean powers are limited to a fraction of a watt, and magnetrons are available to give appreciable peak pulse powers. Power combiners for solid-state sources, or multiple source active array systems would greatly extend the capability of systems if they could be developed. Solid-state noise sources would also be invaluable, especially for radiometric applications. Detector and mixer diodes up to 100 GHz are also now available. Work has been extended to higher frequencies, e.g. to 140 GHz, although here the performance and range of available components are greatly reduced. A frequency of 140 GHz is likely to be beyond the capability of fin-line and dielectric guide, and about the limit for strip line. Rectangular guide and components can be used but at high cost and with poorer performance. Above 140 GHz rectangular waveguide is the only commercially obtainable guide, but cost and performance make it increasingly prohibitive as the frequency is increased. Alternative approaches, such as groove-guide and quasi-optic techniques, will become important as the frequency is increased well above 100 GHz, but relatively little work has been done on such alternatives. Groove-guide could also find application if high powers become available, e.g. through gyratrons which can give mean powers of kilowatts at 300 GHz. The development of non-reciprocal devices in the UK appears to be limited mainly to the GEC Research Centre for this frequency range. Isolators are invariably supplied from the USA. High-power isolators will be needed as source powers increase, and more UK effort seems needed in this general field.

Summing up, the overall impression from the colloquium is that the UK is making significant progress up to 100 GHz in extending lower frequency techniques. Devices, components, circuits and systems are being made by several industrial groups, covering the 95 GHz atmospheric window. Comparatively little has been accomplished however, at higher frequencies. There are significant areas where an extended programme is urgently needed to keep UK technology up to date and fully exploit the potential of this part of the electromagnetic spectrum.

D. J. HARRIS