PHILIPS TECHNICAL REVIEW

VOLUME 39
1980

Philips Technical Review is published by the Research Laboratories of N.V. Philips' Gloeilampenfabrieken, Eindhoven, The Netherlands, and is devoted to the investigations, processes and products of the laboratories and plants which form part of or cooperate with enterprises of the Philips group of companies. In the articles the associated technical problems are treated along with their physical or chemical background. The Review covers a wide range of subjects, each article being intended not only for the specialist in the subject but also for the non-specialist reader with a general technical or scientific training.

The Review appears in three editions: English, Dutch and German, all identical in contents. There are twelve numbers per volume, each of about 32 pages. A yearly index is given for each volume and an index covering ten volumes appears every five years (the latest one was included in Volume 35, 1975).

Editors:	Dr J. W. Broer
	Dipl.-Phys. R. Dockhorn, Editor-in-chief
	Dr E. Fischmann
	Ir J. A. Klaassen
	Dr J. L. Sommerdijk
	Dr J. Ubbink
	W. van der Weel, assistant editor
English edition:	D. A. E. Roberts, B.Sc., M. Inst. P., A.I.L. (Redhill, Surrey)
German edition:	Dipl.-Phys. A. Hoyer (Hamburg)

© N.V. Philips' Gloeilampenfabrieken, Eindhoven, The Netherlands, 1981.
Articles may be reproduced in whole or in part provided that the source 'Philips Technical Review' is mentioned in full; photographs and drawings for this purpose are available on request. The editors would appreciate a complimentary copy.

Subject index, Volume 39, 1980

	Page
Aluminium, electrodeposition of	87
Atomic emission spectrometry (AES), microwave-induced plasma as an excitation source for	65
Bi-aspheric lenses, manufacture with COLATH	229
Camera tubes for television, new concept for	201
Chromatofocusing, a new protein-separation method	125
COLATH, a numerically controlled lathe for very high precision	229
Computer, SAMEN/SAMO in designing a	134
Correction plates for projection television, optical, production of	15
Cryopumps, application in industrial vacuum technology	246
Data communication, mobile-radio, error control in	172
Data transmission, baseband, integrated echo canceller for	102
Deflection coils of the 30AX colour-picture system	154
Deflection coils, television, equipment for measuring the magnetic fields of	277
DKDP crystals for use in the TITUS tube	50
Document handling with Megadoc system	329
Echo canceller, integrated, for baseband data transmission	102
EDDY computer program for calculation and design of electro-mechanical devices	78
Electrodeposition of aluminium	87
Electro-mechanical devices, calculation and design of	78
Electron resists, negative, for VLSI	346
Electroradiography, medical	19
Emission spectrometry, atomic, microwave-induced plasma for	65
Field-effect transistor, low-noise microwave GaAs	269
FM sound broadcasting, station and programme identification in	216
GaAs field-effect transistor, low-noise microwave	269
Glass, transparent single-point turning of	92
Glass fibres, optical, precision fracture of	245
HEIS method for determination of implantation profiles	1
High-pressure sodium lamps, new generation of	211
Institute for Nuclear Physics Research (IKO), 35 years of cooperation with (special issue, pp. 285-328):	
history of IKO	286
cyclotron	290
radioisotopes	294
Geiger-Müller counters	296
semiconductor detectors	298
BOL	302
(continued in right-hand column)	

	Page
(IKO, continued)	
electronics for nuclear physics	312
engineering technology	315
ion-beam technology	320
linear electron accelerators	325
Ion implantation in semiconductors	1
Laser, semiconductor, for information read-out	37
Lathe (COLATH), numerically controlled, for very high precision	229
Magnetic-field measurement of television deflection coils, equipment for	277
Magnetic properties of a microscopic particle, measurement of	48
Medical electroradiography — its potential and limitations	19
Megadoc, a modular system for electronic document handling	329
Microwave GaAs field-effect transistor, low-noise	269
Mobile-radio data communication, error control in	172
Modelling and simulation as an aid in designing a computer	134
Passivation layers, method of measuring mechanical stresses in	130
Pressure gauge, fast, for hot and corrosive gases	344
Projection television, production of optical correction plates for	15
Protein separation by means of chromatofocusing	125
Resonant cavity for MIP in AES	65
SAMEN/SAMO program package for modelling and simulation in designing a computer	134
Satellite television reception, direct, low-noise 12 GHz front-end designs for	257
Semiconductor laser for information read-out	37
SIMS method for determination of implantation profiles	1
Sodium lamps, high-pressure (SON lamps), new generation of	211
Spiral-groove bearings, grease-lubricated	184
Station and programme identification in FM sound broadcasting	216
Strain-gauge transducers, thin-film	94
Television camera tubes, new concept for	201
Television deflection coils, see Deflection coils	
Television reception, direct satellite, low-noise 12 GHz front-end designs for	257
Thin-film strain-gauge transducers	94
TITUS tube, DKDP crystals for use in	50
Transducers, thin-film strain-gauge	94
Turning of glass, transparent single-point	92
Vacuum technology, industrial, application of cryopumps in	246
Video cassette recorder, grease-lubricated spiral-groove bearings in	184
VLSI, negative electron resists for	346

Author index, Volume 39, 1980

Page

Akkerman, H. J.
Engineering technology * 315
Aldefeld, B.
Calculation and design of electro-mechanical devices . . 78
Bakker, P.
Linear electron accelerators * 325
Baudet, P., M. Binet and D. Boccon-Gibod
Low-noise microwave GaAs field-effect transistor . . . 269
Beenakker, C. I. M., P. W. J. M. Boumans and P. J. Rommers
A microwave-induced plasma as an excitation source for atomic emission spectrometry 65
Behr, J.-P., P. Pernards, B. Schendel and J. Schwandt
Modelling and simulation as an aid in designing a computer 134
Belouet, C.
DKDP crystals for use in the TITUS tube 50
Berg, J. F. M. van de, T. E. G. Daenen, G. Krijl and R. E. van de Leest
The electrodeposition of aluminium 87
Bethe, K. and D. Schön
Thin-film strain-gauge transducers 94
Binet, M., see Baudet, P.
Boccon-Gibod, D., see Baudet, P.
Boumans, P. W. J. M., see Beenakker, C. I. M.
Brehm, R., K. van Dun, J. Haisma and J. C. G. Teunissen
Transparent single-point turning of glass 92
Brinkman, G. A.
Radioisotopes * 294
Daenen, T. E. G., see Berg, J. F. M. van de
Dantzig, R. van
BOL * 302
Dun, K. van, see Brehm, R.
Engelen, G. A. J. van, J. L. M. Hagen and W. A. L. Heijnemans
An equipment for measuring the magnetic fields of television deflection coils 277
Finck, J. C. J., H. J. M. van der Laak and J. T. Schrama
A semiconductor laser for information read-out 37
French, R. C. and P. J. Mabey
Error control in mobile-radio data communication . . . 172
Gerwen, P. J. van, W. A. M. Snijders and N. A. M. Verhoeckx
An integrated echo canceller for baseband data transmission 102
Gielis, G. C. M., J. B. H. Peek and J. M. Schmidt
Station and programme identification in FM sound broadcasting 216
Gijsbers, T. G.
COLATH, a numerically controlled lathe for very high precision 229
Hagen, J. L. M., see Engelen, G. A. J. van
Haisma, J., see Brehm, R.
Harrop, P., P. Lesartre and T. H. A. M. Vlek
Low-noise 12 GHz front-end designs for direct satellite television reception 257
Heijnemans, W. A. L., J. A. M. Nieuwendijk and N. G. Vink
The deflection coils of the 30AX colour-picture system . 154
—, see Engelen, G. A. J. van
Hofker, W. K.
Geiger-Müller counters * 296
Semiconductor detectors * 298
Electronics for nuclear physics * 312
Ion-beam technology * 320
— and J. Politiek
Ion implantation in semiconductors 1

Page

Howden, H.
Production of optical correction plates for projection television 15
Jacobs, C. A. J. and J. A. J. M. van Vliet
A new generation of high-pressure sodium lamps . . . 211
Jagt, J. C. and P. W. Whipps
Negative electron resists for VLSI 346
Knowles, J. E.
Measuring the magnetic properties of a microscopic particle 48
Krijl, G., see Berg, J. F. M. van de
Laak, H. J. M. van der, see Finck, J. C. J.
Leest, R. E. van de, see Berg, J. F. M. van de
Lens, G. A., see Zaengel, T.
Lesartre, P., see Harrop, P.
Luijckx, G.
The cyclotron * 290
Mabey, P. J., see French, R. C.
Muijderman, E. A., G. Remmers and L. P. M. Tielemans
Grease-lubricated spiral-groove bearings 184
Nie, A. G. van
A method of measuring mechanical stresses in passivation layers 130
Nieuwendijk, J. A. M., see Heijnemans, W. A. L.
Peek, J. B. H., see Gielis, G. C. M.
Pernards, P., see Behr, J.-P.
Peschmann, K. R.
Medical electroradiography — its potential and limitations 19
Politiek, J., see Hofker, W. K.
Remmers, G., see Muijderman, E. A.
Rommers, P. J., see Beenakker, C. I. M.
Roosmalen, J. H. T. van
A new concept for television camera tubes 201
Scheer, J. J. and J. Visser
Application of cryopumps in industrial vacuum technology 246
Schendel, B., see Behr, J.-P.
Schmidt, J. M., see Gielis, G. C. M.
Schön, D., see Bethe, K.
Schrama, J. T., see Finck, J. C. J.
Schwandt, J., see Behr, J.-P.
Sluyterman, L. A. Æ.
Chromatofocusing, a new protein-separation method . . 125
Snijders, W. A. M., see Gerwen, P. J. van
Teunissen, J. C. G., see Brehm, R.
Tielemans, L. P. M., see Muijderman, E. A.
Verhoeckx, N. A. M., see Gerwen, P. J. van
Vink, N. G., see Heijnemans, W. A. L.
Visser, J., see Scheer, J. J.
Vlek, T. H. A. M., see Harrop, P.
Vliet, J. A. J. M. van, see Jacobs, C. A. J.
Vos, J. A. de
Megadoc, a modular system for electronic document handling 329
Waalwijk, J. M. and N. Wiedenhof
The Institute for Nuclear Physics Research 'has finished its work' * 286
Whipps, P. W., see Jagt, J. C.
Wiedenhof, N., see Waalwijk, J. M.
Zaengel, T. and G. A. Lens
Fast pressure gauge for hot and corrosive gases 344

* Article in the special issue 'Thirty-five years of cooperation with IKO'.

PHILIPS TECHNICAL REVIEW

VOLUME 39, 1980, No. 1

Ion implantation in semiconductors

W. K. Hofker and J. Politiek

The breathtaking developments in LSI technology in the last few years would have been impossible without the progress made in the doping of semiconductors by ion implantation. At Philips the fundamental research on this method is mainly carried out in the Amsterdam department of the Research Laboratories, while most of the applications research is carried out at Philips Research Laboratories at Redhill, in England. In the article below, the authors describe the principles of the method and discuss some of the experiments that have laid the basis for many practical applications.

Introduction

When a solid is bombarded with energetic ions, most of the ions penetrate into the solid and come to a stop at a certain depth: they become 'implanted' there. Ion bombardment changes the chemical composition and the atomic lattice of the substrate ('implantation damage') and therefore alters its chemical and physical properties. In this article we shall be mainly concerned with the changes produced in the band structure (energy levels) as a result of the ion bombardment, which is important in producing *N*- or *P*-type electrical conductivity in semiconductors. In this process the substrate is bombarded with ions that act as donors or acceptors when they have become established at a lattice site. Since the damage that the bombardment causes to the crystal lattice is generally a disadvantage for practical applications, the bombardment is normally followed by a thermal treatment (annealing), which repairs the damage completely or in part.

Until recently thermal diffusion [1] was the method most widely used for doping semiconductors. Compared with this method ion implantation has a number of important advantages. The doping can be much more accurate, even when very low doping is required. It is also easy to adjust the penetration depth of the ions with this method; complicated depth profiles can be produced in a fairly simple way by varying the energy and dose of the ions. Another advantage is that the dose can be very evenly distributed in the lateral direction. An important advantage is that almost any element can be implanted. The method can also be used for driving the ions through thin surface layers. On the other hand, if a masking layer is to be used, it does not have to be an oxide layer, as in the thermal-diffusion method, but can be an evaporated metal film, a photoresist, etc.

There are however some disadvantages: an annealing treatment is usually necessary to repair the damage to the crystal lattice, the equipment is complicated and therefore expensive, and the method is not suitable for deep implantations. Limitations are then set by the ion energy required and by the energy dissipation in the substrate.

In such a case the ion-implantation method is often combined with the thermal-diffusion method. That is to say the ions are first implanted at a lower energy and subsequently 'driven in' to the required depth by means of thermal diffusion (drive-in diffusion). This means, of course, that some of the advantages of ion implantation are lost.

Dr Ir W. K. Hofker and Dr Ir J. Politiek are with Philips Research Laboratories, Amsterdam Department.

[1] See for example A. Schmitz, Philips tech. Rev. **27**, 192, 1966.

In some cases a method is used in which the element to be implanted is first deposited in the form of a layer on the substrate surface (e.g. by evaporation), and the surface is then bombarded with ions such as argon or krypton. With this method, known as recoil implantation, it is possible to produce a very thin doped layer that has a high doping concentration at the surface [2].

The potential of ion implantation was first recognized in 1952 at the Bell laboratories [3], but applications of significance were not initially found. It was not until the 1960s, when efforts were made to find a better doping method for the development of such devices as semiconductor radiation detectors [4] and high-frequency MOS transistors [5], that ion implantation came fully into its own.

The work at Philips Research Laboratories on this method was started in 1962 in the department located in Amsterdam. In a later phase cooperation began with Philips Research Laboratories at Redhill, England, where particular attention was being paid to the development of various applications. The more fundamental, physical aspects were still studied in Amsterdam. The results of this study, in so far as they relate to research on semiconducting silicon, are the subject of this article.

Before discussing these results we shall describe the implantation equipment, the measures adopted to ensure a uniform and reproducible dose, the collision processes that play a part in implantation, the theoretical principles that make it possible to calculate the penetration depth, etc., and two of the methods that can be used to measure the doping profiles.

The practical value of ion implantation is already very considerable. The present advances made in the development of LSI technology, for example, would not have been possible without the use of ion implantation [6].

Equipment and procedure

Equipment

Fig. 1 shows a schematic diagram of one of the types of implantation equipment used in our research. The ions are extracted from the source *So* by an electrostatic field, and are then focused on the mass-defining slit *Sl* by means of an electrostatic lens *E* and an analysing magnet *An*. The field in this magnet is adjusted so that only the desired ions pass through the slit *Sl* to the output; the undesired ions are intercepted. The ions are then accelerated in an 'accelerating tube' *Ac* and focused by the lens *F* on to the substrate *S*, which is contained in the target chamber *Ch*. The *X* and *Y* scan plates sweep the ion beam in a zigzag pattern (for example) over both the substrate and the

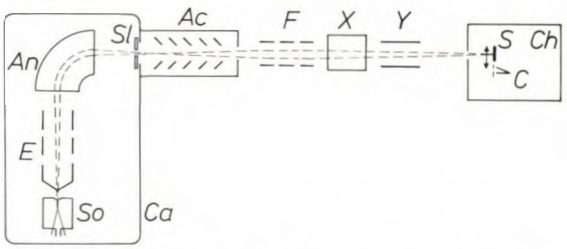


Fig. 1. Diagram of an ion-implantation system. *Ca* high-voltage cage. *So* ion source. *E* and *F* electrostatic focusing lenses. *An* analysing magnet. *Sl* slit. *Ac* accelerating tube. *X* and *Y* scan plates. *Ch* target chamber. *S* substrate. *C* Faraday cage. *Ac, F, X* and *Y* form the beam line.

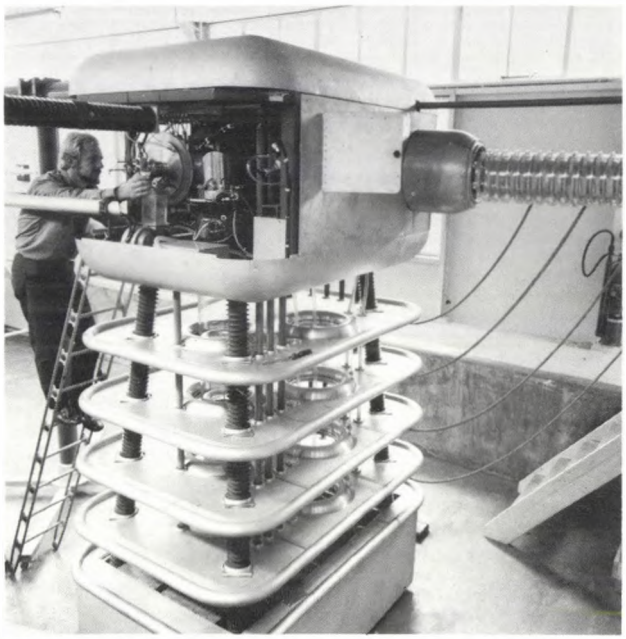


Fig. 2. View of a high-voltage cage, containing the ion source and analysing magnet. A research worker is fitting an ion source. The accelerating tube can be seen on the right.

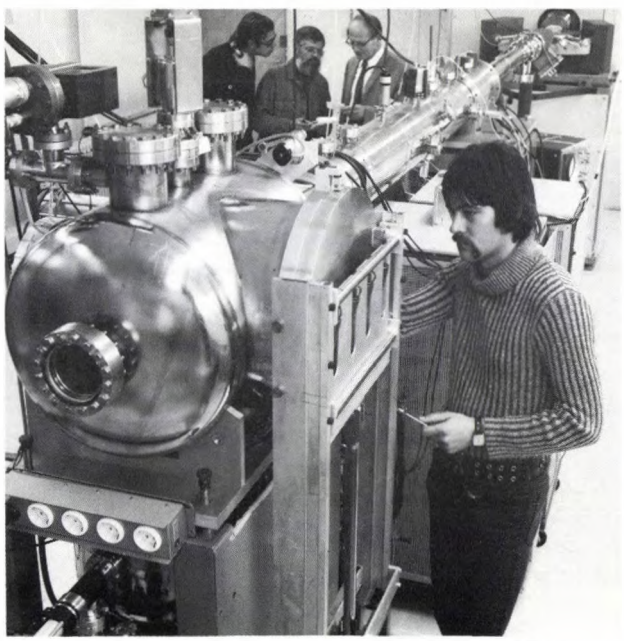


Fig. 3. Target chamber (left foreground) with part of the beam line behind it.

Faraday cage *C* situated next to it. The charge received in this cage, integrated over the implantation time, is a measure of the dose per unit area received by the substrate. The ion beam scans the substrate for as long as necessary to reach the required dose.

Fig. 2 is a photograph of the high-voltage cage (which contains the ion source and the analysing magnet) of a 500 kV accelerator that we use for research purposes. *Fig. 3* shows the beam tube and the target chamber of the same equipment.

Fig. 4 shows the target chamber of an apparatus used for implanting ions with very high beam currents. In this target chamber a number of substrates are mounted on a rotating disc. The periodic interruption of the bombardment of the separate substrates by the rotation is advantageous for the energy dissipation in each substrate. This makes it possible to implant large doses with an ion-beam current high enough for the total implantation time to remain within acceptable limits, without the substrate becoming too hot.

The disc rotates at a speed of about 1 rev/s and at the same time makes a radial movement at a speed inversely proportional to the distance between the beam and the centre of the disc. This corrects for the fact that the tangential velocity at which the substrates pass the beam varies with the distance to the centre of the disc.

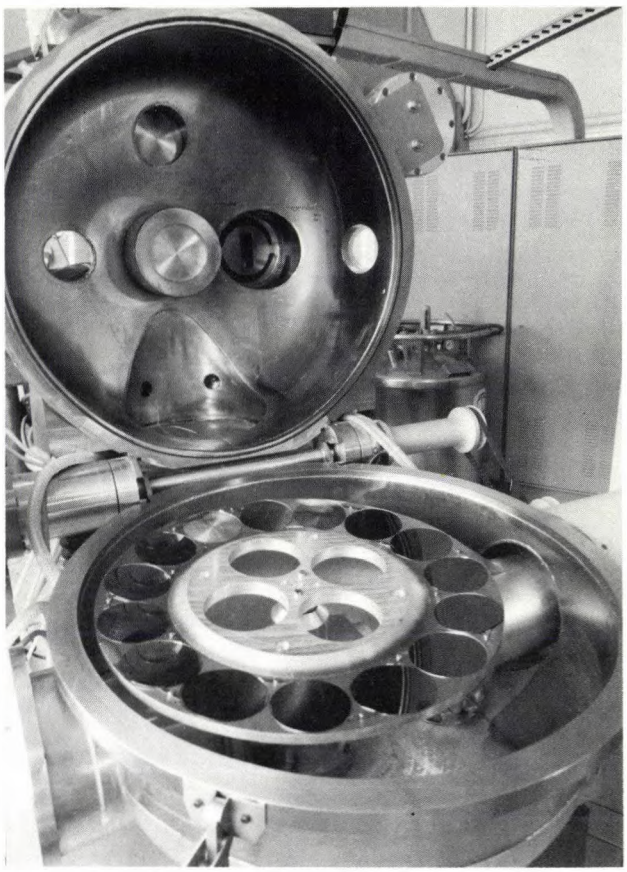


Fig. 4. Target chamber (opened) of a system for high beam currents. The specimens are mounted on a disc, which rotates during the bombardment; the disc is then in the vertical position.

Uniformity and reproducibility of the dose

When the direction of the ion beam coincides with an open crystallographic direction or channel (e.g. a ⟨100⟩, a ⟨110⟩ or a ⟨111⟩ direction), the ions penetrate much further than when the bombardment is in a random direction. This situation is usually found in perpendicular bombardment of the substrate; the fact that the ion beam is moved over the substrate during such a bombardment has the effect of producing marked variations in the penetration depth of the implanted ions. To avoid this effect the substrate is bombarded not vertically but obliquely (*fig. 5a*). At the centre of the substrate the angle of incidence is usually 7° to 11° from the normal. When the substrate is scanned with the ion beam in this situation, in the pattern shown in fig. 5*b*, there is a change in both the angle of incidence and the distance from the point of incidence on the substrate to the centre of the scan plates (*X* or *Y*). Both changes cause variations in the

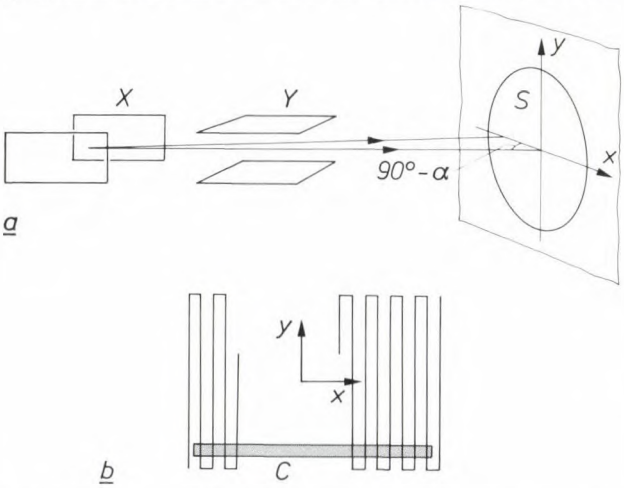


Fig. 5. Diagrams of our method of ion bombardment for obtaining a uniformly distributed dose. *a*) The ion beam strikes the centre of the substrate at an angle of $90° - \alpha$ to the surface. (The angle α is generally between 7° and 11°.) *b*) The pattern described by the ion beam, deflected by the scan plates *X* and *Y*, on the substrate, and the Faraday cage *C*. More details are given in the text.

[2] See for example J. M. Shannon, Appl. Phys. Lett. **24**, 369, 1974, and **25**, 75, 1974.
[3] R. S. Ohl, Bell Syst. tech. J. **31**, 104, 1952.
[4] See for example W. K. Hofker, Philips tech. Rev. **27**, 323, 1966, and W. K. Hofker, K. Nienhuis and J. C. Post, Philips tech. Rev. **30**, 13, 1969, in particular footnote [3].
[5] See J. M. Shannon, Philips tech. Rev. **31**, 267, 1970.
[6] See for example H. Bosma and W. G. Gelling, Philips tech. Rev. **37**, 267, 1977.
Other applications investigated in Philips Research Laboratories where ion implantation is not used as a doping method are discussed in:
G. A. Acket, J. J. Daniele, W. Nijman, R. P. Tijburg and P. J. de Waard, Philips tech. Rev. **36**, 190, 1976;
W. J. M. J. Josquin, Proc. Int. Conf. on Ion beam modification of materials, Budapest 1978, p. 1433;
D. P. Oosthoek, J. A. den Boer and W. K. Hofker, Proc. Eur. Conf. on Ion implantation, Reading 1970, p. 88.

dose distribution over the surface of the scan. Calculations show that it is necessary to modify the control voltage, particularly for the x-direction, in such a way as to compensate for this dose variation.

Deviations in dose uniformity are then mainly caused by beam-current fluctuations during the scanning of the substrate. The uniformity can be appreciably improved by implanting the dose in a number of scans or sweeps (e.g. 50). *Fig. 6* gives an idea of the dose uniformity that can be obtained in this way.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
0	0	0	0	0	0	589	598	598	599	599	599	599	599	599	599	599	599	599	600	598	0	0	0	0	0	0
0	0	0	0	0	599	582	599	599	597	600	599	599	600	599	599	598	600	599	600	598	0	0	0	0	0	0
0	0	0	0	579	601	599	599	599	600	600	599	599	0	0	599	596	601	599	600	604	600	601	0	0	0	0
0	0	0	0	598	599	599	599	600	600	600	600	600	600	599	600	600	600	600	600	600	600	600	0	0	0	0
0	0	0	599	598	599	600	600	600	600	600	603	599	600	600	600	600	601	601	601	601	601	601	601	0	0	0
0	0	584	599	599	600	600	600	601	600	600	600	600	600	600	600	601	602	600	602	602	601	601	601	602	0	0
0	0	599	599	600	601	600	601	601	601	600	601	601	600	601	601	601	601	601	602	600	602	602	602	602	0	0
0	0	598	624	601	601	602	601	602	601	602	601	601	602	601	601	601	602	601	601	601	602	602	602	602	0	0
0	599	599	600	601	602	602	602	601	602	602	601	602	602	601	602	601	602	600	601	601	602	602	602	603	602	0
0	559	599	601	601	602	601	602	602	602	602	603	602	601	602	603	602	602	601	602	601	602	603	602	602	603	0
0	598	0	600	600	601	601	602	603	602	602	603	602	602	602	603	602	602	602	602	602	602	602	603	604	603	0
599	599	600	600	600	601	600	601	603	602	602	603	602	602	602	603	603	603	602	602	602	602	603	604	604	604	0
599	600	600	601	601	600	601	602	602	601	602	603	602	602	602	603	603	603	603	603	602	602	603	603	603	604	605
600	600	601	601	601	601	601	601	601	601	602	601	600	600	602	603	603	604	603	603	603	603	603	604	605	604	605
600	600	601	600	601	601	601	601	601	601	601	600	600	600	598	600	602	604	603	604	603	604	603	605	604	604	606
601	577	601	602	601	601	602	602	601	602	601	601	601	600	600	600	601	602	603	604	604	604	604	605	604	605	606
600	600	0	0	602	602	602	602	602	602	602	601	599	0	0	600	601	602	602	603	603	603	604	604	0	0	606
600	601	602	603	602	604	603	603	603	603	601	600	601	601	600	601	601	601	602	604	602	603	604	604	605	605	605
601	602	602	603	603	603	603	604	603	603	602	602	602	603	601	602	602	602	603	602	602	603	603	603	604	604	605
598	601	602	604	603	603	604	603	602	604	603	603	603	603	602	602	602	603	602	603	603	603	603	603	603	603	560
494	602	603	603	603	604	603	603	603	604	603	603	603	604	603	604	603	603	603	603	603	603	604	604	603	602	397
599	602	602	603	0	0	604	604	604	604	603	604	603	604	604	604	594	604	603	603	603	603	603	604	604	599	370
0	0	603	603	603	603	603	603	603	603	603	604	604	605	605	604	603	605	604	604	605	604	604	604	603	598	373
0	603	603	603	603	604	603	603	603	604	603	604	604	605	604	605	605	605	604	605	604	604	597	604	604	507	0
0	603	603	603	604	604	602	604	603	603	604	604	604	604	604	605	604	605	605	605	605	606	605	602	604	327	0
0	602	604	603	603	603	604	604	602	604	603	603	604	605	604	604	604	605	605	605	606	606	604	606	605	379	0
0	0	603	603	604	604	603	603	603	604	604	603	604	605	603	605	604	605	605	605	606	605	606	606	605	0	0
0	0	603	603	604	603	603	603	603	603	604	604	603	604	604	605	605	605	605	606	606	606	606	605	607	0	0
0	0	0	603	604	603	603	604	603	603	603	604	604	605	604	605	605	606	606	605	606	606	605	605	0	0	0
0	0	0	603	603	604	603	603	603	603	603	604	604	605	605	605	605	605	606	606	606	605	605	0	0	0	0
0	0	0	0	604	602	604	604	603	0	604	604	604	0	0	605	605	605	606	604	604	605	605	0	0	0	0
0	0	0	0	0	602	603	606	603	604	604	605	604	605	604	604	605	606	605	605	605	0	0	0	0	0	0
0	0	0	0	0	0	603	602	603	604	604	603	604	605	605	604	606	605	605	605	604	0	0	0	0	0	0
0	0	0	0	0	0	0	602	603	602	604	604	603	604	604	605	604	605	605	605	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	602	603	603	604	604	604	602	605	0	0	0	0	0	0	0	0	0

Fig. 6. Illustrating the uniformity of dose that can be obtained with the implantation method in fig. 5. Topographical distribution of the sheet resistance in Ω/□ over a substrate after implantation of 10^{14} B^+ ions/cm² at 70 keV. The resistivity at the sites indicated was measured by the Van der Pauw method[7].

The bombardment of the substrate is stopped when the Faraday cage has measured the required dose, but the beam sweep is completed. This has the result that the required dose is always slightly exceeded. To avoid this effect, which adversely influences the reproducibility of the dosage, the implanted dose is measured with the Faraday cage after each complete sweep of the substrate, and a control unit checks whether the implanted dose is too large or too small after each sweep. The next dose is then adjusted appropriately, increasing or decreasing the beam scanning rate.

What happens in the implantation process?

Calculations with the LSS model

The various quantities encountered in the implantation process, such as the energy loss per unit pathlength, the penetration depth, etc., are usually calculated from a model based on a now classical theory due to Niels Bohr [8] and developed by J. Lindhard, M. Scharff and H. E. Schiøtt [9], known as the LSS model.

In this model the implanted ions lose their energy both through collisions with the nuclei of the substrate atoms and through electron interactions. The collision processes between the ions to be implanted and the nuclei of the substrate atoms, which are partly screened by electrons, are determined by Coulomb forces. The collisions here are elastic collisions, i.e. the transferred energy is purely kinetic. This elastic collision process is responsible for the displacement of atoms in the substrate, and hence for the damage produced in it. The interactions of the ions to be implanted with the electrons in the substrate are as-

sociated with excitation and ionization processes, and are therefore inelastic. The energy loss of the implanted ion per unit pathlength can be calculated from the LSS model as a function of the energy (*fig. 7*).

It can be seen from this figure that for nuclear collisions with increasing energy ε the energy loss per unit pathlength passes through a maximum (at ε_1), whereas the energy loss for electronic collisions increases monotonically. Initially the loss is smaller, but at higher energy (higher than ε_2) it becomes greater than the energy loss for nuclear collisions. Thus, whereas the energy absorption is largely due to inelastic collision processes for $\varepsilon > \varepsilon_2$, the elastic collisions predominate for $\varepsilon < \varepsilon_2$. This means that an ion implanted with an energy greater than ε_2 dissipates its energy at the beginning of its penetration into the substrate mainly in collisions with electrons, and that as soon as its energy has decreased to less than ε_2 the energy loss due to nuclear collisions predominates. Since, as already noted, displacements of atoms in the crystal lattice are caused by nuclear collisions, ε_2 is an important quantity.

For implantations in silicon ε_2 is 17 keV for B^+ ions, 140 keV for P^+ ions, 800 keV for As^+ ions and 2000 keV for Sb^+ ions. At an implantation energy of, say, 200 keV the number of substrate atoms displaced on bombardment with As^+ ions is thus greater than for bombardment with B^+ ions. Broadly speaking, the rule is that the higher the atomic number Z of an ion, the greater the number of atoms displaced by the ion.

The LSS model has also been used to calculate the mean penetration depth and the mean spread of the penetration depth for different kinds of ions and different substrate materials [10]. For B^+ ions of 200 keV in silicon the mean penetration depth is 530 nm with a mean spread of 92 nm, and for Sb^+ ions of the same energy the corresponding figures are about 80 and 24 nm.

Transition from crystalline to amorphous state

If the concentration of the atoms displaced by one ion is high, there is a local transformation in the substrate from the single-crystal to an amorphous state. If the implanted ion dose is large enough, the amorphous regions made by individual ions overlap each other, resulting in a continuous amorphous layer.

Whether the substrate becomes amorphous or not depends not only on the dose and the atomic number of the ions used, but also on the temperature of the substrate during bombardment — and hence on the dose rate as well. This is because the lattice can still recover to some extent until the substrate material becomes amorphous, and the rate of recovery increases with the temperature.

Limitations of the LSS model

An assumption of the LSS model is that the substrate atoms are randomly distributed. It is found to be difficult to modify the theoretical model to obtain a better approximation to the actual crystalline nature of the crystal lattice. This is why it is difficult to calculate the 'tails' that occur in implantation profiles for bombardment in a random direction, since these tails are caused by ions that enter the lattice in an open crystallographic direction during the bombardment, directly or after a number of collisions, and are thus able to penetrate further into the crystal than is the case in other directions. A fairly good agreement between theory and experiment is only found for implantations in amorphous or polycrystalline materials [11].

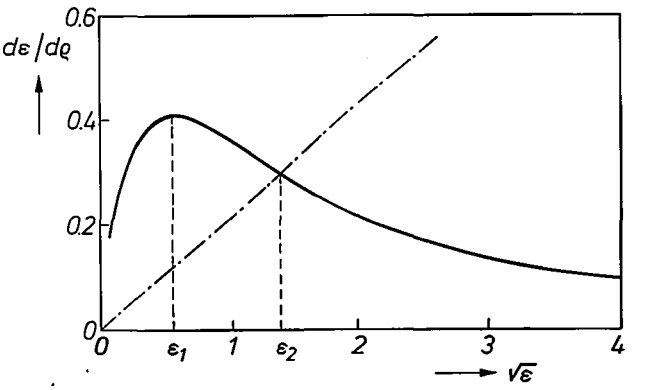


Fig. 7. The calculated energy loss per unit pathlength for ions implanted in a substrate, as a result of collisions with substrate atoms (*solid curve*) and with electrons (*chain-dotted line*), both as a function of the energy of the ions[9]. To obtain a result valid for all elements, 'reduced' values have been used for the range, ϱ and for the energy, ε. The slope of the line giving the energy loss due to collisions with electrons is also a function of the atomic number Z of the ion and that of the substrate atoms. The line shown here refers to boron implantation ($Z = 5$) in silicon ($Z = 14$).

Determination of implantation profiles

Various methods are available for determining the doping profiles produced by ion implantations. Two of these, HEIS and SIMS, will be described here and their particular merits discussed.

HEIS

In the analysis of implantation profiles by the method known as HEIS (*H*igh-*E*nergy *I*on *S*cattering) a beam of light high-energy ions is directed on to a substrate, and the energy distribution of the ions that

[7] L. J. van der Pauw, Philips tech. Rev. **20**, 220, 1958/59; the principle is also explained in: H. B. G. Casimir, Philips tech. Rev. **24**, 341, 1962/63, particularly p. 351.
[8] N. Bohr, Mat.-fys. Medd. Kgl. Dan. Vid. Selsk. **18**, No. 8, 1948.
[9] J. Lindhard, M. Scharff and H. E. Schiøtt, Mat.-fys. Medd. Kgl. Dan. Vid. Selsk. 33, No. 14, 1963.
[10] J. F. Gibbons, W. S. Johnson and S. W. Mylroie, Projected range statistics, 2nd edition, Halsted Press, 1975.
[11] See W. K. Hofker, Philips Res. Repts. Suppl. 1975, No. 8.

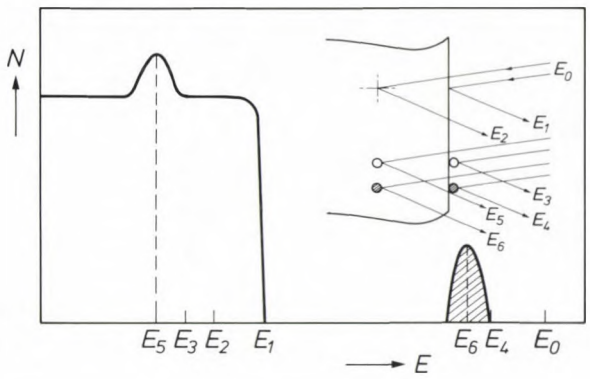


Fig. 8. When a substrate is bombarded with light ions of high energy, the implantation profile in the substrate can be derived from the energy distribution of the ions scattered in a particular direction. A spectrum obtained in this way by the HEIS method is represented schematically, giving a current N of scattered ions as a function of their energy E. *Inset:* various possible types of ion scattering: scattering of ions by *substrate atoms* at the surface (the initial energy E_0 is reduced to E_1) and in the bulk of the lattice (the energy is reduced to E_2), and scattering by *foreign atoms* at the surface (E_3 for light foreign atoms, E_4 for heavy ones) and in the bulk (E_5 for light foreign atoms, E_6 for heavy ones).

ion and the scattering angle. Since light ions are used, the analyses can be carried out to a considerable depth, and the damage resulting from the analysis is slight.

Fig. 8 shows a schematic HEIS spectrum of a silicon substrate containing foreign atoms, some with a mass greater than that of silicon (shown hatched) and some with a smaller mass.

We shall now take a closer look at the shape of such an HEIS spectrum. We shall first discuss the spectrum produced by the ions that are scattered by the substrate atoms (fig. 8, inset). The maximum energy with which ions are scattered by these atoms is E_1, as the result of a collision with a surface atom. If an ion penetrates to a certain depth in the substrate before undergoing a collision, it loses some energy on its way through the substrate as a result of collisions with free electrons. The energy retained by the ion in this case is smaller than E_1 (e.g. E_2). Scattering of this kind can of course take place at any depth, so that a continuous

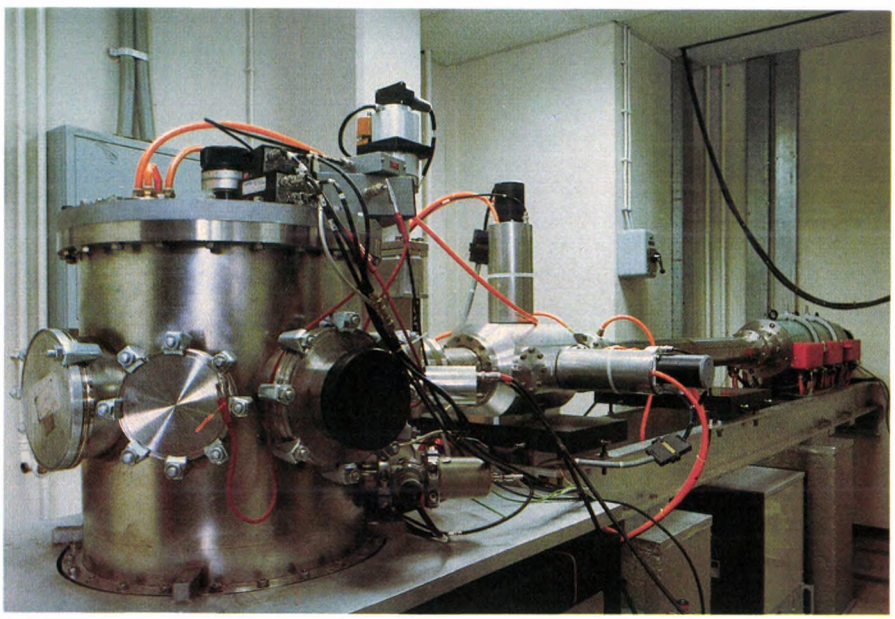


Fig. 9. View of a Van de Graaff accelerator. The photograph on the left shows the chamber in which specimens are placed (left foreground), and part of the beam line, passing through a radiation

are scattered within a particular angle to the incident beam is measured by a semiconductor detector [12]. When high-energy ions are used, interactions with electrons predominate (see fig. 7), and the probability of a collision with an atomic nucleus is small. It may therefore be assumed that the scattered ions that are detected have only collided once with the nucleus of a substrate atom. For this simplest case, the single collision, the mass and as a rule the depth of the scattering atom can be derived from the energy of the scattered spectrum is produced with a maximum energy E_1.

If there are foreign atoms on the surface of the substrate whose mass is smaller than that of the substrate atoms, the energy after scattering from these atoms will be lower than E_1, with the result that its spectrum appears superimposed on the continuous substrate spectrum. If on the other hand the mass of the foreign atoms is greater than that of the substrate atoms, a peak will appear at an energy greater than E_1 (hatched).

If the foreign atoms are not on the substrate but *in* it (at the same depth), there will be a shift to lower energy values, e.g. from E_3 to E_5 and from E_4 to E_6, owing to the additional energy loss of the ions. A distribution of impurities in the bulk, as obtained upon implantation, will be indicated by a broadening of the HEIS spectrum.

Since, at the high energies used, the energy loss of the ions fired into the substrate to a depth where the implanted ions are situated is almost entirely determined by interactions with electrons and is virtually independent of the energy of the scattered ions, the energy scale can be replaced by a depth scale, so that the implantation profile can be read directly from the spectrum. As the energy of the scattered ions also depends on the mass of the scattering ion, the depth scale does however differ for each individual element.

The fast light ions required are generated in our laboratory by a Van de Graaff accelerator, shown in *fig. 9*. With this accelerator we can produce, for example, protons and He^+ ions with an energy of 2.5 MeV. By way of illustration, *fig. 10* shows the spectrum, measured by the method described, of a silicon substrate in which arsenic has been implanted.

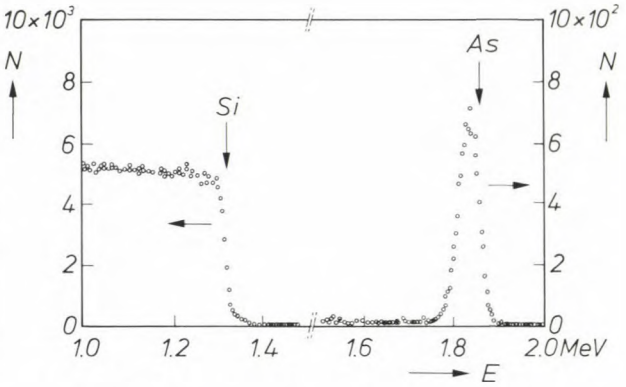
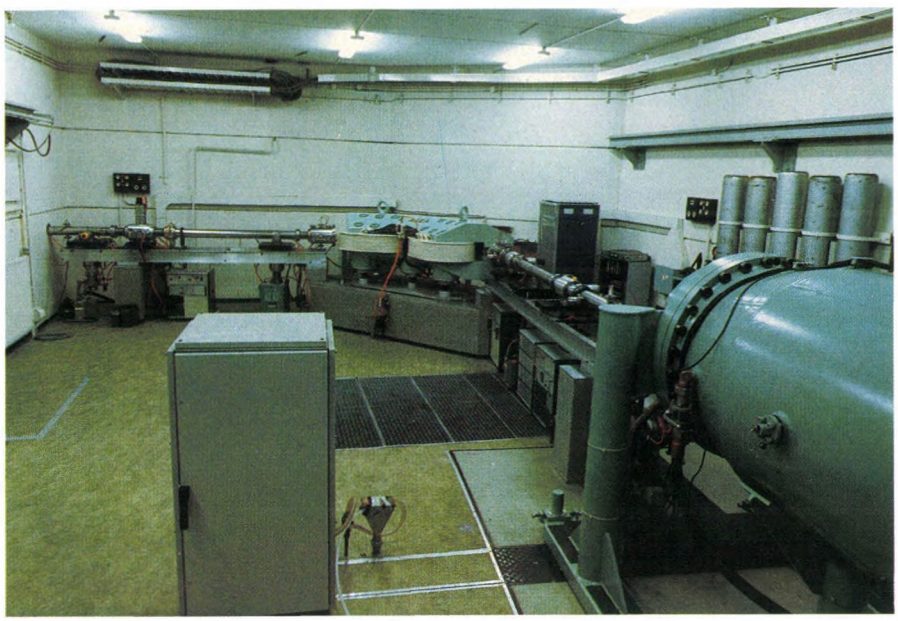


Fig. 10. HEIS spectrum of an arsenic-implanted silicon substrate. The substrate was bombarded with He^+ ions of energy 2.3 MeV for the analysis. The arsenic was implanted at 100 keV; the dose was 8.5×10^{15} ions/cm². The arrows indicate the energy values of the He^+ ions that have been scattered by Si and As atoms *at the surface*.



shield. The part of the beam line on the other side of the radiation shield can be seen in the right-hand photograph; the generator is at the extreme right.

An interesting feature of HEIS is that it makes it possible to determine whether or not implanted ions are located at lattice sites. This is done by comparing the spectra obtained by bombardment in a 'dense' direction with that obtained by bombardment in an open (low-index) crystallographic direction. In a dense direction, as distinct from an open direction, the ion beam has a chance of hitting *every* atom. The conclusions that may be drawn from the difference between these two spectra are illustrated in *fig. 11*. The figure shows that an ion not located at a lattice site (ion *A*) is 'seen' both by an ion beam entering in a dense direction and by a beam in an open direction. On the other hand, an ion at a lattice site (ion *B*) is *not* seen by a beam entering in an open direction.

Fig. 12 shows as an example two HEIS spectra of a silicon substrate bombarded with arsenic ions. The

[12] See for example W. F. van der Weg and Y. Tamminga, Acta Electronica **19**, 47, 1976.

circles denote the recorded spectrum resulting from a beam of He^+ ions in a dense direction, and the triangles denote the spectrum obtained on bombardment in an open direction in silicon, in this case a ⟨100⟩ direction. From the large difference between these two As spectra it is possible to calculate that about 40% of the As ions are not located at lattice sites.

It will be evident from the foregoing that the method is suitable for the study of the damage caused by implantation. *Fig. 13* shows HEIS spectra of a silicon substrate bombarded with N^+ ions, where the analysing ion beam was directed both in a dense direction (circles) and in a ⟨111⟩ direction (triangles). The conclusion to be drawn from these spectra is that the silicon of the substrate has become amorphous to a depth of 210 nm.

SIMS

As noted above, in HEIS measurements the spectrum of the foreign atoms whose mass is smaller than that of the substrate atoms coincides with that of the substrate atoms. This adversely affects the detection limit of these elements. A method that has been found to be better in such a case is SIMS (*S*econdary-*I*on *M*ass *S*pectrometry). In this method the substrate is bombarded with low-energy ions in such a way that atomic layers are 'excavated' one after the other from the substrate surface by sputtering [13]. The sputtered particles — which need not only be atoms or ions, but can also be clusters of atoms or ions — are passed through a mass analyser. The current produced by the ions or ion clusters at the output of the detection system of the analyser is a measure of the concentration of the element being investigated in the various sputtered layers. By determining the thickness of the sputtered layer this current as a function of time can be converted into the concentration profile as a function of depth. This technique has been found very useful for studying implantations such as that of boron in silicon [11].

Some experimental results

In this section we shall discuss a number of experimental results that have provided some insight into the effects that occur when ion implantation is used in semiconductor technology. We shall confine ourselves mainly to studies of boron and antimony implantations in silicon. Boron is an acceptor and gives *P*-type conduction, whereas antimony is a donor. Another interesting difference between the atoms ^{11}B and ^{121}Sb is their great difference in mass, which causes considerable differences in penetration depth and in the damage produced.

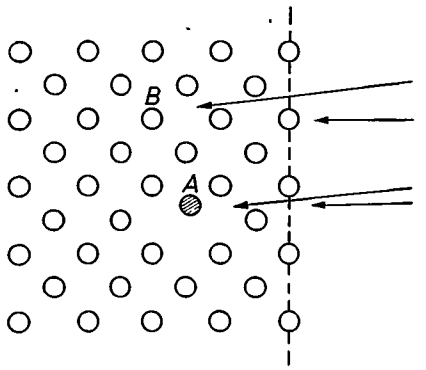


Fig. 11. Showing how HEIS can demonstrate the presence of atoms at non-substitutional sites. The non-substitutional atom *A* is 'seen' both by ions in a random direction and by ions that have entered in an open crystallographic direction or channel. The atom *B*, on the other hand, is not detected by ions entering a channel.

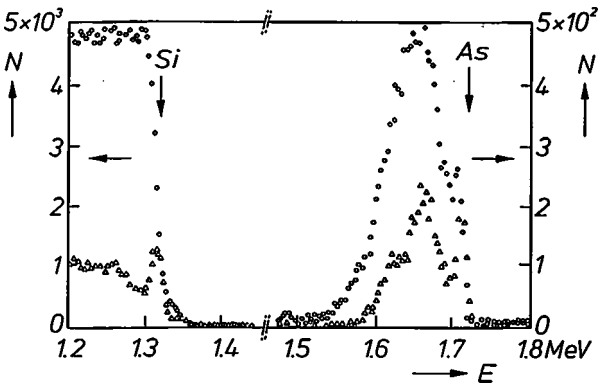


Fig. 12. HEIS spectra obtained by bombarding a silicon wafer (He^+ ions with an energy of 2 MeV) in a random direction (*circles*) and in a ⟨100⟩ direction. Arsenic (dose 10^{16} ions/cm²; energy 100 keV) was implanted in the wafer, and the wafer was then annealed for 30 minutes at 900 °C. The difference in the areas of the arsenic peaks indicates that 40% of the arsenic is located at non-substitutional sites.

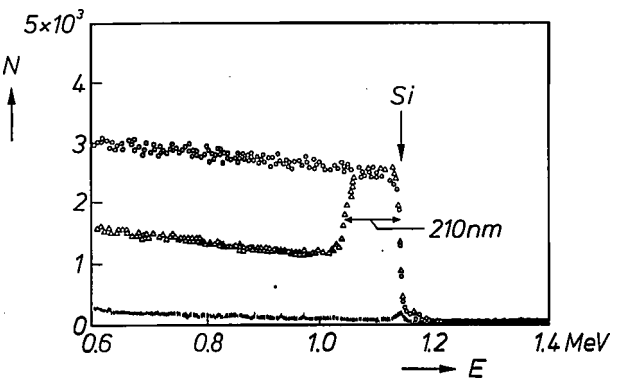


Fig. 13. HEIS spectra obtained by bombarding silicon (He^+ ions with an energy of 2.0 MeV) in a random direction (○) and in a ⟨111⟩ direction (△). N^+ ions were implanted in the substrate with a dose of 10^{16} ions/cm² and an energy of 70 keV. The lower spectrum, shown for comparison, was obtained by bombarding undoped silicon in a ⟨111⟩ direction. The region where the upper two curves coincide corresponds to a zone of the substrate where no distinction between the two directions mentioned above can be observed, and which must therefore be regarded as amorphous.

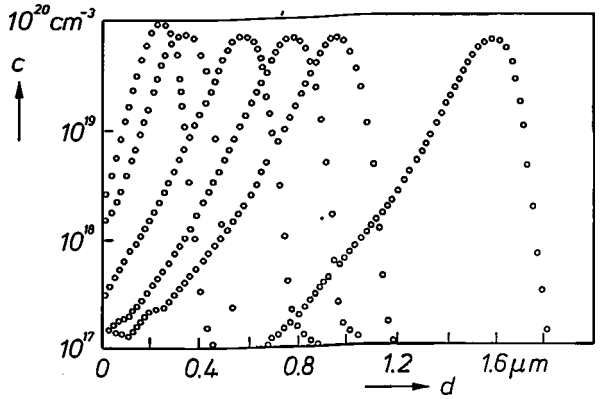


Fig. 14. Depth distributions, determined by SIMS, of boron (concentration c as a function of depth *d*) implanted in polycrystalline silicon with a dose of 10^{15} ions/cm^2 and energies of 70, 100, 200, 300, 400 and 800 keV.

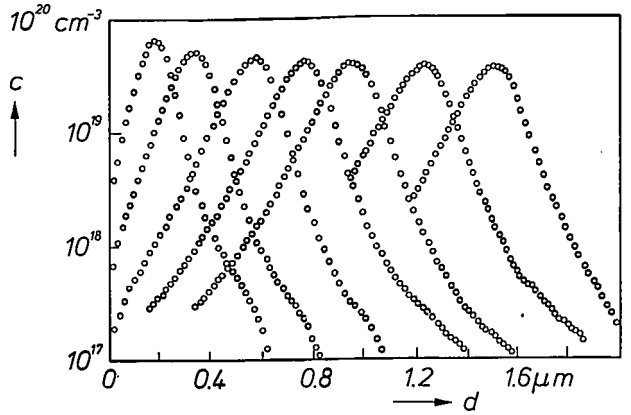


Fig. 15. Depth distributions of boron implanted in crystalline silicon in a ⟨763⟩ direction — a 'dense' crystallographic direction — with a dose of 10^{15} ions/cm^2 and energies of 50, 100, 200, 300, 400, 600 and 800 keV. 'Tails' occur because of the channelling effect.

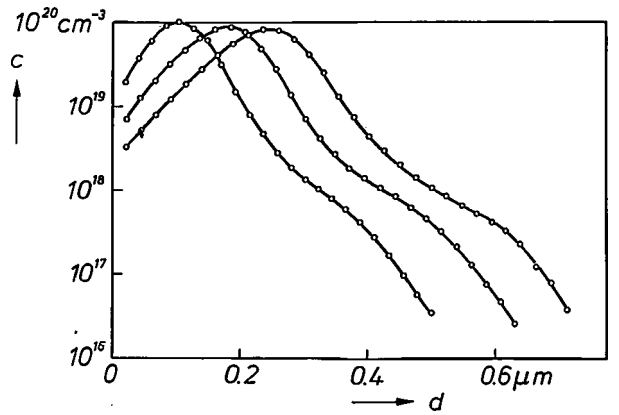


Fig. 16. Depth distributions of boron implanted in crystalline silicon in a ⟨763⟩ direction with a dose of 10^{15} ions/cm^2 and a relatively low energy (30, 50 and 70 keV). The influence of the channelling effect is even more pronounced here than in the case of fig. 15.

Studies of the process of implanting boron atoms, which have a smaller mass than silicon atoms, have mainly been made by SIMS, whereas antimony studies have largely been made by HEIS.

We shall begin by discussing a number of characteristic depth profiles obtained by ion implantation, and the way in which they are affected by heat treatment (annealing). We shall then consider the process of recovery from implantation damage and the changes in electrical behaviour that are connected with this process.

Implantation profiles before and after annealing

We shall first examine the shape of the implantation profiles obtained when no thermal treatment is applied. *Fig. 14* shows profiles obtained by implantations with mono-energetic boron ions in very fine *polycrystalline* silicon. These profiles have a gently sloping 'leading edge' and a steep 'trailing edge', which is due to the domination of interactions between the boron ions and the electrons in the substrate at the beginning of the path travelled by the ions, and the predominance of collisions with the lattice atoms at the end of the path. This applies to implantation energies that are relatively high compared with ε_2 (fig. 7). The experimental and calculated profiles show good agreement in this case [11].

Fig. 15 shows a series of profiles obtained by implantation in *single-crystal* silicon. To minimize channelling effects the implantation was performed in a dense crystallographic direction, a ⟨763⟩ direction. Comparison with the preceding figures shows that the trailing edge is less steep and has a slight tail. In low-energy implantations there is even a distinct hump in this region (*fig. 16*). The tails are caused by the channelling effect. This has been very clearly demonstrated for phosphorus implantations in silicon [14].

By varying the bombardment angle the channelling effect can be enhanced. *Fig. 17* shows a boron distribution for an implantation in a ⟨110⟩ direction, which is the most open crystallographic direction in silicon. For comparison the figure also shows the distribution for an implantation, with the same energy, in a dense direction, again a ⟨763⟩ direction. The ⟨110⟩ profile has two maxima. The left-hand maximum is due in the first place to boron ions that are scattered by collisions with atoms at the head of a row immediately they enter the crystal, and in the second place to boron ions that can get no further at a later stage because the channels become blocked. The maximum

[13] M. Croset, Rev. tech. Thomson-CSF 3, 19, 1971.
W. K. Hofker, H. W. Werner, D. P. Oosthoek and H. A. M. de Grefte, Radiation Effects 17, 83, 1973.
[14] P. Blood, G. Dearnaley and M. A. Wilkins, J. appl. Phys. 45, 5123, 1974.

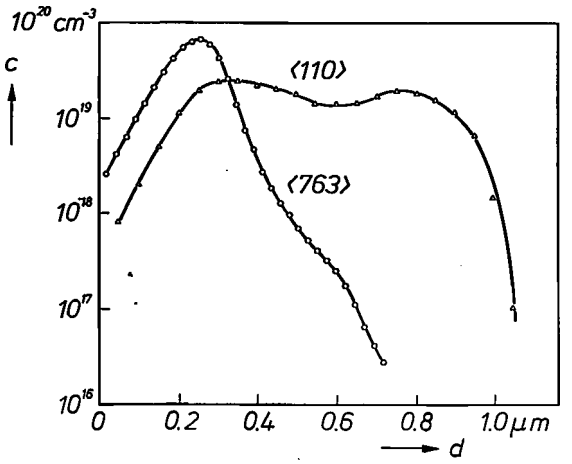


Fig. 17. Depth distributions of boron implanted in crystalline silicon in the directions ⟨763⟩ and ⟨110⟩ (an 'open' channel). Dose 10^{15} ions/cm^2, energy 70 keV. The channelling effect is now very clear.

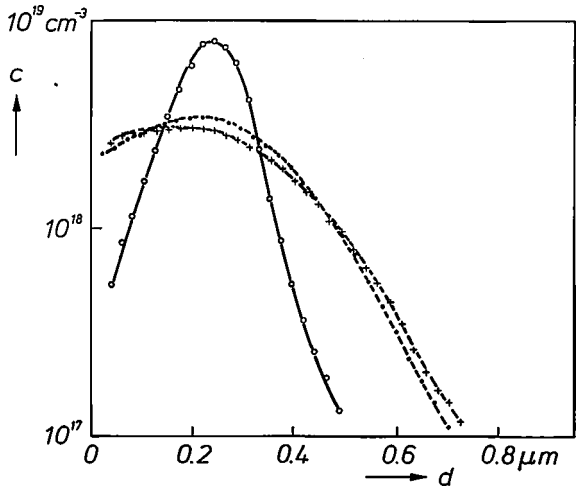


Fig. 18. Depth distributions, determined by SIMS, of boron implanted in silicon (dose 10^{14} ions/cm^2, energy 70 keV). The distributions were measured before (○) and after one anneal (●) of 35 minutes at 1000 °C, and after two anneals for the same time and temperature (+).

on the right is formed by ions that have entered through a channel.

Let us now see how the implantation profile changes on annealing. We shall consider for this purpose the effect of annealing on the diffusion process. *Fig. 18* shows the result of two successive anneals after a boron implantation in single-crystal silicon. It appears from the resultant profiles that the diffusion process is appreciably slower during the second anneal than during the first. We believe that this slowing down can be explained from the fact that after implantation a substantial part of the boron is located at non-substitutional sites, so that in the first phase of the anneal a non-substitutional diffusion predominates. This stops when the boron has settled at lattice sites, e.g. on being trapped by vacancies. The measurements indicate that a substitutional diffusion process then prevails, which is about ten times slower.

Annealing gives different profiles if the solid solubility is exceeded during the implantation. In such a case there is precipitation of the implanted atoms. *Fig. 19* shows profiles obtained immediately after a boron implantation and after one, two and three anneals of 35 minutes at 1000 °C. It can be seen that the boron is 'immobile' where the concentration is highest. The immobility indicates that the boron has precipitated. This assumption is supported by *fig. 20*, which gives the depth distribution of both the boron and the charge carriers. The immobile boron fraction is seen to be electrically inactive, as would be expected for a precipitate.

The effect of an anneal on the profiles is also influenced by the damage caused to the crystal lattice by

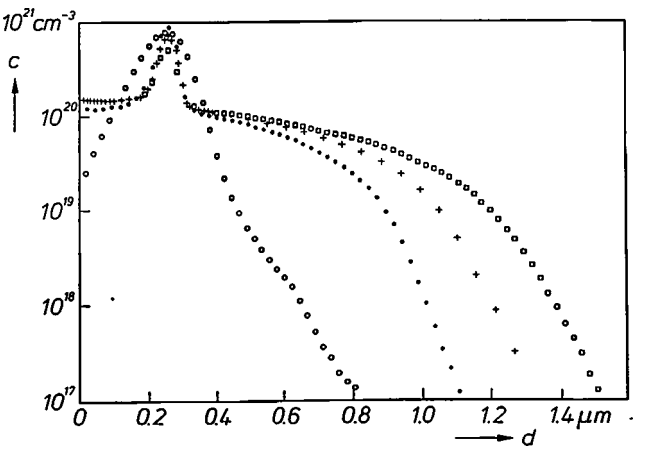


Fig. 19. Depth distributions of boron implanted in silicon (dose 10^{16} ions/cm^2, energy 70 keV). The distributions were measured before (○) and after one (●), two (+) and three (□) anneals of 35 minutes at 1000 °C. The peak occurring in all the distributions is attributed to precipitated boron.

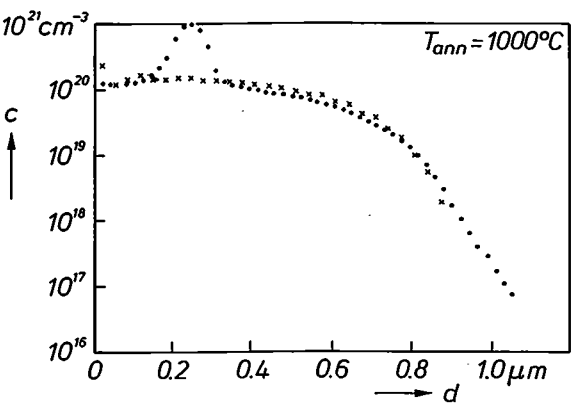


Fig. 20. Depth distribution of boron-implanted silicon (●) and the corresponding distribution of charge carriers (×). Dose 10^{16} ions/cm^2, energy 70 keV. The substrate was annealed at 1000 °C for 35 minutes. This result confirms the assumption made in the caption to fig. 19.

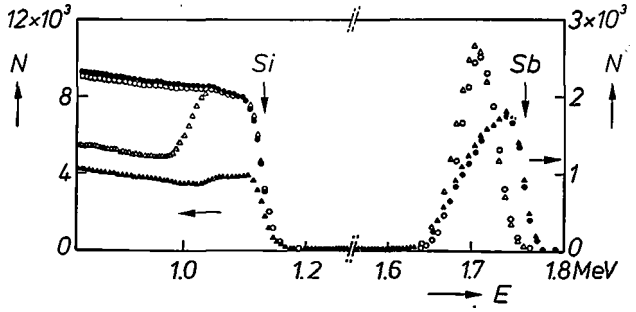


Fig. 21. Spectra obtained with HEIS (He^+ ions, 2 MeV) from ⟨111⟩ silicon doped with antimony (dose 10^{16} ions/cm², energy 150 keV). The diagram shows spectra measured immediately after implantation by bombardment in a random direction (○) and in a ⟨111⟩ direction (△), and spectra measured after annealing (for 40 minutes at 1000 °C) by bombardment in a random direction (●) and in a ⟨111⟩ direction (▲). It is clear that the anneal has caused antimony to diffuse to the surface.

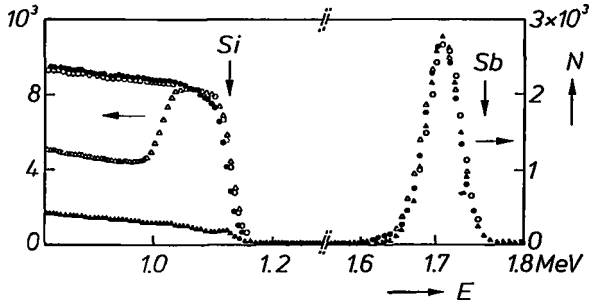


Fig. 22. Spectra like those in fig. 21, but now for silicon with a ⟨100⟩ instead of a ⟨111⟩ orientation. Under these conditions there is no diffusion of antimony to the surface.

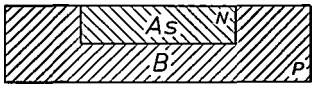


Fig. 23. Diagram of a *P-N* junction that can be produced in silicon by implantation of boron and arsenic.

the implantation. This is particularly clear from experiments with antimony implantations.

Fig. 21 gives HEIS spectra of an antimony-implanted silicon substrate with ⟨111⟩ crystal orientation, before and after annealing. This figure shows that annealing has made the antimony diffuse to the crystal surface. No such effect is found when the same implantation is performed by a silicon substrate with ⟨100⟩ orientation (*fig. 22*). A comparison of the two figures reveals that after the anneal considerably more damage remains in the ⟨111⟩ substrate than in the ⟨100⟩ substrate. This is connected with the observed diffusion of the antimony to the surface of the ⟨111⟩ substrate [15], since it may be explained as thermal diffusion accelerated by the implantation damage.

We should also note that during an anneal above 1000 °C the antimony that has precipitated in this way at the substrate surface diffuses back into the substrate by means of a substitutional diffusion process.

In boron implantations the crystal damage is appreciably less than in antimony implantations, at least in normal doses, so that the thermal diffusion to the surface described above is not so pronounced.

We shall now consider the case of an implantation of both boron and arsenic ions in silicon [16].

This combination is often used for producing *P-N* junctions (*fig. 23*). In a combined implantation of this type the damage caused by the heavier ions (arsenic) can affect the diffusion of the lighter ions. *Fig. 24* shows boron distributions obtained after first implanting boron, annealing at 1000 °C, then implanting arsenic and finally annealing at various temperatures. *Fig. 25* shows the boron distributions produced when the anneal between implantations was omitted.

In both cases the distributions contain irregularities. We believe that the irregularities found at a depth of about 0.2 μm are due to mechanical stresses set up by the implantation [11]. At the location where an amorphous layer has formed, as a result of the arsenic implantation, the mechanical stresses have been relieved by plastic deformation. Immediately outside this region, however, where the silicon is still only just crystalline, the mechanical stresses are at a maximum. This depth corresponds to the depth of the irregularities in the boron distributions as shown in figs 24 and 25.

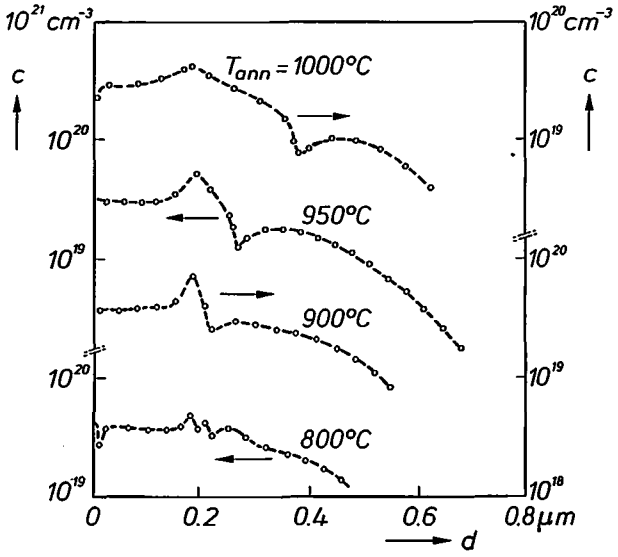


Fig. 24. Depth distributions, determined by the SIMS method, of boron in silicon that has been implanted with boron (10^{15} ions/cm², 70 keV) and with arsenic (10^{16} ions/cm², 150 keV) and which has been annealed for 35 minutes at the temperatures indicated. Between the boron and the arsenic implantation the silicon was also annealed for 35 minutes at 1000 °C.

[15] W. J. M. J. Josquin and Y. Tamminga, Appl. Phys. 15, 73, 1978.

We have established that this effect can be avoided by first implanting the arsenic and then annealing to allow the lattice to recover before implanting the boron.

The irregularities encountered at a depth greater than 0.2 μm — at 0.27 μm in the curves in figs 24 and 25 for T_{ann} = 950 °C and at 0.4 μm in the curves for T_{ann} = 1000 °C — are electrical in origin. As others have found, these depths correspond to the location of the *P-N* junction, which moves to a greater depth in the substrate at higher values of T_{ann} [17]. At the *P-N* junction an electric field has an effect on the concentration of the boron ions.

Implantation damage and electrical behaviour

The annealing is intended not only to allow the lattice to recover from the damage caused by ion bombardment, but also to electrically activate the implanted ions. The ions cannot perform the donor or acceptor functions required of them until they have occupied a lattice site in the restored lattice.

As can be seen in *fig. 26*, the recovery of a layer that has become amorphous due to ion bombardment takes place from the part of the substrate underneath it that has remained undamaged. Investigations carried out elsewhere have shown that the recovery rate depends closely on the orientation of the bombarded region [18]. For example, the recovery rate in silicon is about ten times faster for a ⟨100⟩ orientation than for a ⟨111⟩ orientation.

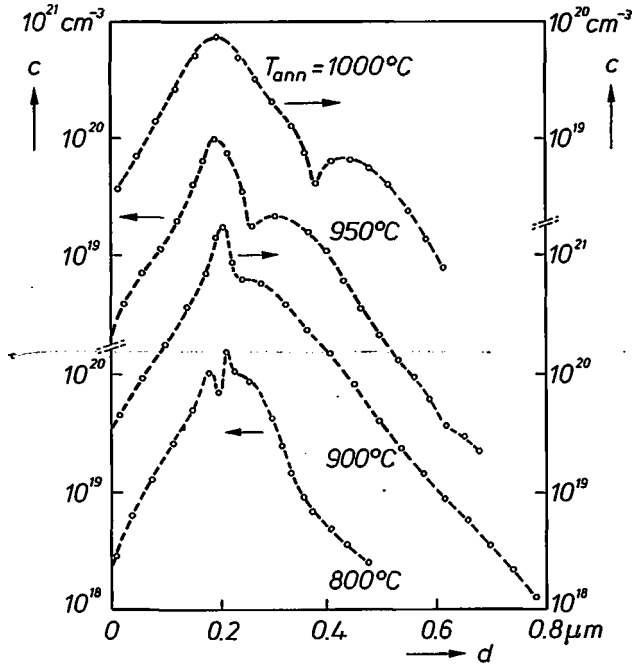


Fig. 25. As in fig. 24, but without the anneal between the boron and arsenic implantations. The irregularities in the boron distributions, seen both in this figure and in fig. 24, are discussed in the text.

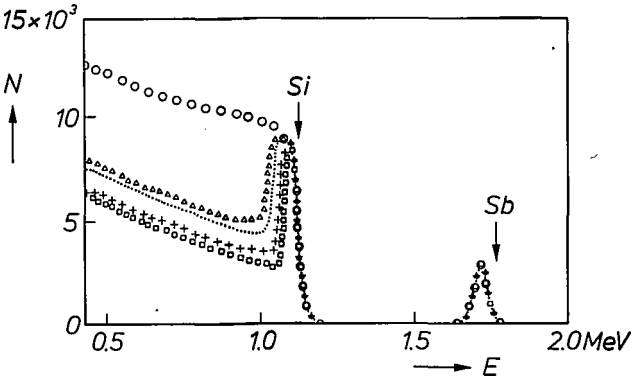


Fig. 26. HEIS spectra (He^+ ions, 2 MeV) of silicon with a ⟨100⟩ orientation, bombarded with antimony (dose 10^{16} ions/cm^2, energy 150 keV). The diagram shows the spectra resulting from implantation in a random direction (○) and in a ⟨111⟩ direction, measured both immediately after implantation (△) and after annealing the substrate at 500 °C for 20 minutes (●), 40 minutes (+) and 170 minutes (□). It can clearly be seen that the layer that has become amorphous because of bombardment gradually recovers during the anneal; the recovery starts *in the substrate* and travels in the direction of the surface.

Marked differences are also found in the ultimate effect of the annealing for the orientations mentioned above, as can be seen from *figs 27* and *28*. These are photomicrographs, made with a transmission electron microscope, of ⟨100⟩ and ⟨111⟩ substrates, both bombarded with the same antimony dose at the same energy, and both subjected to the same anneal. In the ⟨111⟩ substrate there is still considerable residual damage to be seen, whereas in the ⟨100⟩ substrate the restored crystalline structure predominates (see also the comments to figs 21 and 22). Longer annealing does not remove this difference.

The residual damage in the ⟨111⟩ substrate sometimes contains line dislocations (not visible here). We have found that the occurrence of such line dislocations depends on the temperature of the substrate during implantation [19]. It has also been found that oxidation processes that take place after the implantation may help to cause line dislocations [20]. The formation of these line dislocations is particularly undesirable in the manufacture of transistors and integrated circuits. If such locations cross the area of

[16] W. K. Hofker, H. W. Werner, D. P. Oosthoek and N. J. Koeman, in: Applications of ion beams to materials 1975, Proc. Conf. Coventry (Inst. Phys. Conf. Ser. No. 28, 1976), p. 13.
[17] J. F. Ziegler, G. W. Cole and J. E. E. Baglin, Appl. Phys. Lett. **21**, 177, 1972.
[18] L. Csepregi, E. F. Kennedy, J. W. Mayer and T. W. Sigmon, J. appl. Phys. **49**, 3906, 1978.
[19] W. K. Hofker, W. J. M. J. Josquin, D. P. Oosthoek and J. R. M. Gijsbers, Proc. Int. Conf. on Ion beam modification of materials, Budapest 1978, p. 1121.
[20] P. Ashburn, C. Bull, K. H. Nicholas and G. R. Booker, Solid-State Electronics **20**, 731, 1977.
C. Bull, P. Ashburn, G. R. Booker and K. H. Nicholas, Solid-State Electronics **22**, 95, 1979.

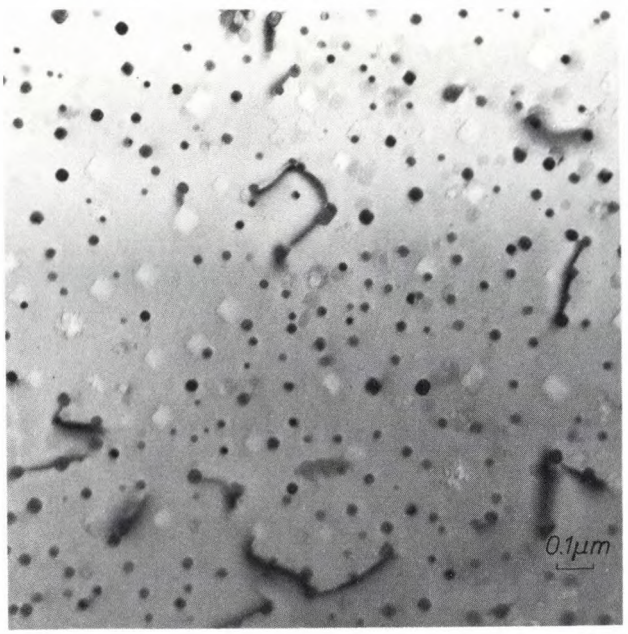


Fig. 27. Photomicrograph, made by transmission electron microscope, of a ⟨100⟩ silicon substrate doped with antimony (dose 10^{16} ions/cm^2, energy 150 keV), and annealed for eight hours at 1130 °C.

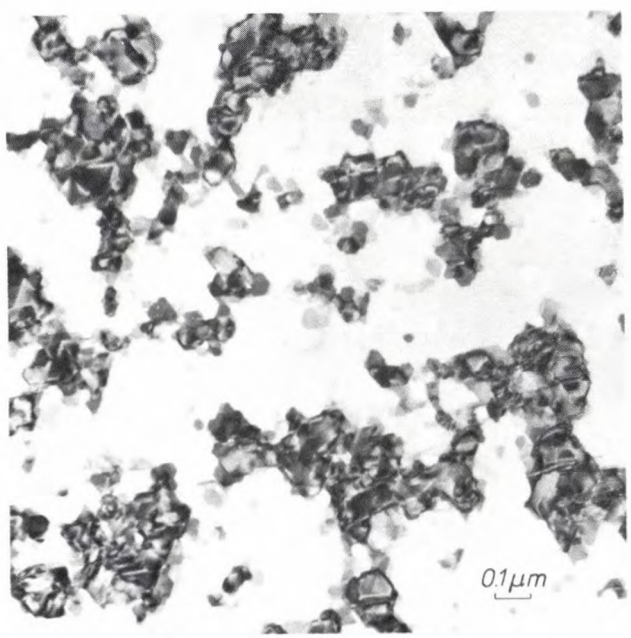


Fig. 28. As fig. 27, but for a silicon substrate with a ⟨111⟩ orientation.

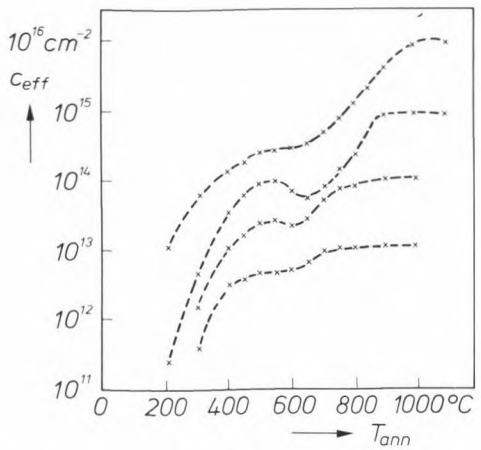


Fig. 29. Effective surface concentration c_{eff} of charge carriers in a silicon substrate bombarded with boron at an energy of 70 keV with doses of 10^{13}, 10^{14}, 10^{15} and 10^{16} ions/cm^2 as a function of the temperature during 35 minutes of annealing. The temperature was raised by 50 °C at each step.

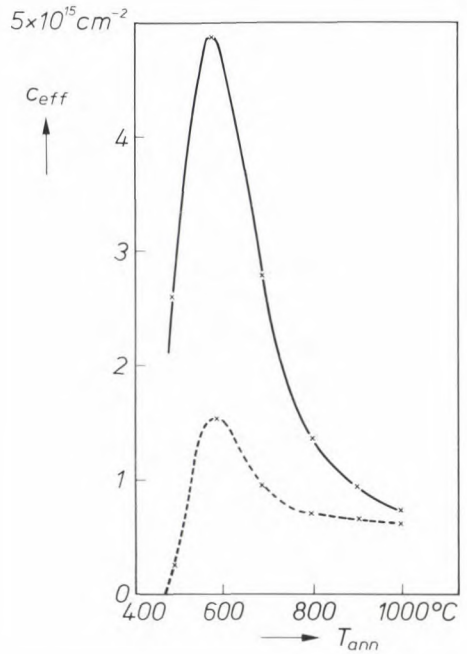


Fig. 30. Effective surface concentration c_{eff} of charge carriers in a silicon substrate with a ⟨100⟩ orientation (*solid curve*) and a ⟨111⟩ orientation, again as a function of the annealing temperature. The substrate was bombarded with 10^{16} antimony ions/cm^2 at an energy of 150 keV.

the *P-N* junction they have an adverse effect on the electrical behaviour of the junction.

The solid solubility of the implanted ions also plays an important part in their activation, as we have already seen in figs 19 and 20. *Fig. 29* shows that the electrical activation of boron ions implanted at a dose of 10^{13} ions/cm^2 is completed after an anneal of 40 minutes at a temperature of 800 °C, whereas at a dose of 10^{16} ions/cm^2 a temperature of 1100 °C is required for the same annealing time. This difference in behaviour arises because a quantity of boron precipitates at the higher implantation doses, since the solid solubility limits are exceeded, and this precipitate only goes into solution at a higher temperature or after longer annealing.

Fig. 29 reveals another and surprising effect: at a temperature of about 500 °C a further increase in temperature produces no further increase, or only a very

small one, in the electrical activity. Sometimes it may even decrease (reverse annealing). This effect is possibly a result of interactions between point defects (which arise during the recovery from the implantation damage) with substitutional boron atoms, which thus become electrically inactive.

In our laboratory we have also observed an anomalous *increase* in the electrical activity, for implantations with large doses of antimony [16]. This effect is particularly pronounced in implantations in silicon with a ⟨100⟩ orientation (*fig. 30*), where it is found that the antimony concentration at lattice sites is considerably in excess of the solid solubility. After an anneal at 600 °C this substitutional antimony concentration can increase to $9\times10^{20}/cm^3$, although the solubility is no more than about $7\times10^{19}/cm^3$. This result, which is in good agreement with electrical measurements, is of course a purely kinetic effect, the result of the particularly high preference that antimony atoms show for lattice sites. After an anneal at higher temperature (or a longer anneal at 600 °C) the electrical activity consequently decreases again owing to the precipitation of antimony that then occurs.

The lattice damage caused by implantations can have a marked effect on the reverse current of *P-N* junctions. This is particularly the case if damage has been caused — and is not removed by annealing — in the depletion layer of the junction. The lattice damage itself and the impurities trapped in the damaged region, such as Cu and Au, can produce energy levels about half-way into the forbidden band, which act as generation/recombination centres and cause an increase in the reverse current. This does not usually happen for damage outside the depletion layer, which may even be beneficial if it is able to remove impurities from the depletion layer by gettering [21].

One indication that the residual damage determines the reverse current to a great extent can be obtained from the magnitude of the reverse current measured on substrates that have undergone the same treatment and differ only in crystallographic orientation. As already mentioned (page 12), the residual damage resulting from implantation in substrates with a ⟨100⟩ orientation is less than that in substrates with a ⟨111⟩ orientation. This agrees with the finding that the reverse current in the first case can easily be an order of magnitude smaller.

Summary. The results are described of physical research on ion implantation in semiconductors, carried out mainly in the Amsterdam department of Philips Research Laboratories. Also discussed are the advantages and disadvantages of ion implantation as against thermal diffusion, the implantation equipment, a theoretical model of the implantation process, and two methods used for determining implantation profiles, HEIS and SIMS.

[21] P. Zandveld, Solid-State Electronics **21**, 721, 1978.

Production of optical correction plates for projection television

H. Howden

Ever since the advent of television there has been interest in the idea of generating a large picture from a relatively small cathode-ray tube by optical projection on to a screen. In the early years, before glass technology was far enough advanced to permit the mass production of large picture tubes, this technique was envisaged as being suitable for home use [1]. Later, however, it became possible to produce such large picture tubes that this application for projection television ceased to be of interest. Nevertheless, it is still necessary for some professional purposes [2] to have a picture size (of the order of 1 m^2 or larger) that considerably exceeds the maximum dimensions of a television screen.

The professional mammoth TV projector [3] shown in *fig. 1* was developed in response to this requirement. The optical section of this apparatus consists of three barrels that project the red, green and blue basic pictures (one from each barrel) on top of one another on the viewing screen. Each barrel is fitted with a cathode-ray tube that acts both as an object and as a light source. In addition each barrel contains a concave spherical mirror and an aspheric Schmidt plate for the correction of spherical aberration; see *fig. 2*.

Fig. 1. The Philips professional mammoth TV projector. This is fitted with projection tubes with a diameter of 125 mm, a mirror with a diameter of 400 mm and a correction plate of 280 mm. The diagonal of the projected picture can be given any value between 1.8 and 7.5 m.

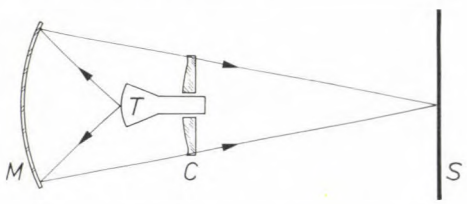


Fig. 2. The optical section of the mammoth TV projector shown in fig. 1. *T* cathode-ray tube with a red, green or blue fluorescent screen. *M* spherical concave mirror. *C* aspheric Schmidt correction plate. *S* projection screen.

H. Howden, C.Eng., M.I.Mech.E., is with Philips Research Laboratories, Redhill, Surrey, England.

[1] P. M. van Alphen and H. Rinia, Philips tech. Rev. **10**, 69, 1948/49.

[2] An application for medical teaching has been described by W. A. Holm and F. H. J. van der Poel, Philips tech. Rev. **20**, 327, 1958/59.

[3] T. Poorter and F. W. de Vrijer, Philips tech. Rev. **19**, 338, 1957/58.

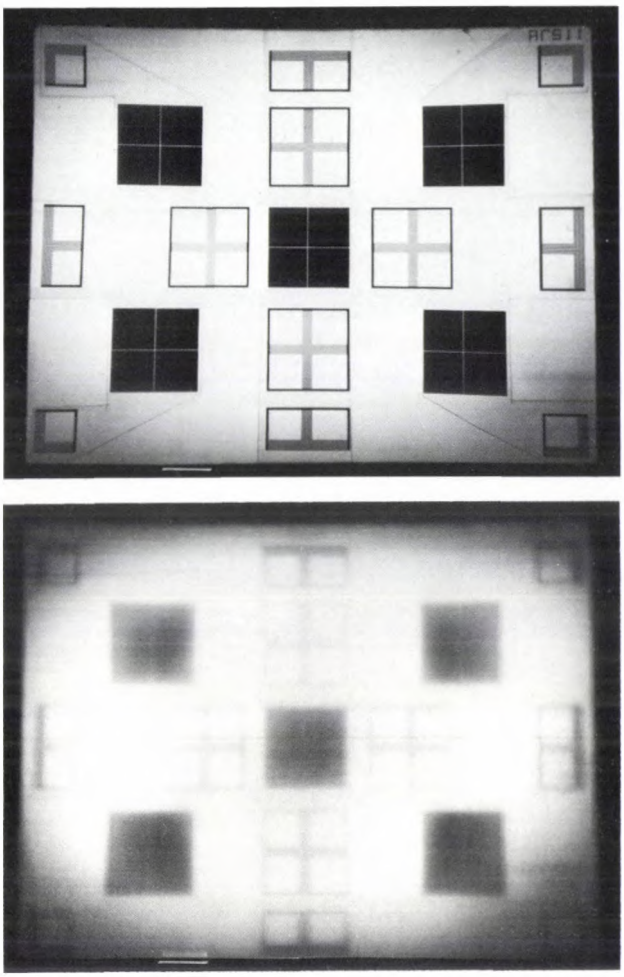


Fig. 3. A test pattern projected by a system similar to that in fig. 2 (*a*) *with* correction plate and (*b*) *without* correction plate. The effect of spherical aberration is clearly visible in (*b*).

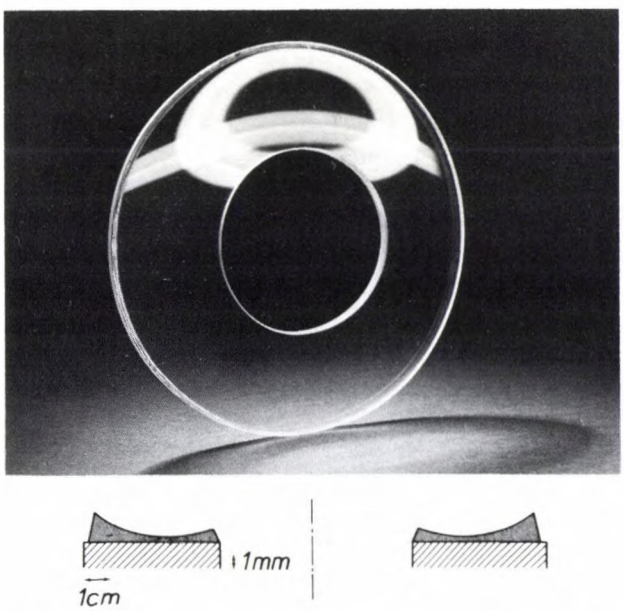


Fig. 4. Photograph of a correction plate of cast synthetic resin. The diagram below is a schematic cross-section of the correction plate (*grey:* the resin layer; *hatched:* the glass substrate). The strongly curved reflection of the fluorescent lamps in the room comes from the profiled resin surface. The other reflection is caused by the back of the glass substrate.

The essential importance of this correction plate for the quality of the picture projected by such an optical system is illustrated in *fig. 3*.

Because of the complicated profile, the manufacture of the correction plate — originally developed by Bernard Schmidt for optical systems for photographing the stars — led to problems right from the start. Schmidt himself used a flat circular glass plate, which he placed on the rim of a circular vessel. He evacuated this vessel until the glass plate reached a certain amount of curvature, and then he ground and polished the plate to a calculated spherical profile. After removal from the vessel the glass plate had approximately the desired profile on the ground side.

A method used later at Philips was based on the use of gelatin, which was poured into a mould in solution, then left to set, removed from the mould, dried and hardened [4]. The correction plates of the mammoth TV projector shown in fig. 1 were made of polymethyl methacrylate (Perspex); the desired profile was formed on a precision lathe with the aid of a master template.

With the benefit of the experience acquired with the mammoth TV projector Philips have for some time been working on the development of the optics and tubes for a smaller television projector. This projector can be used in semi-professional activities, e.g. in medical teaching, schools, hotels, exhibition rooms and at airports, etc. These applications involve large numbers and this led to a search for a more appropriate process for producing correction plates [5].

In the process that we have developed the Schmidt correction plate is *cast* with the aid of a profiled glass master mould. Epoxy resin has been used for the casting material, which is permanently applied to a glass plate. The correction plate is shown in *fig. 4*.

A numerically controlled milling machine was used to grind the desired profile in the glass master mould. The profile was polished on a specially designed machine. The accuracies were ± 6 μm for grinding and ± 1 μm for polishing.

The most suitable casting material was found to be a mixture of two different liquid resins made by CIBA-Geigy. Both constituents belong to the group of epoxy resins, which, in comparison with polyester, silicone, urethane, polystyrene and acrylic resins, were found to have the most appropriate properties for the particular combination of requirements

[4] H. Rinia and P. M. van Alphen, Philips tech. Rev. **9**, 349, 1947/48.

[5] The cathode-ray tubes have a screen 75 mm in diameter. The brightness is such that a peak luminance of the order of 100 cd/m² is obtained on a selectively reflecting projection screen of area about 1 m²; this is comparable with cinema performance. The contrast is 30:1.

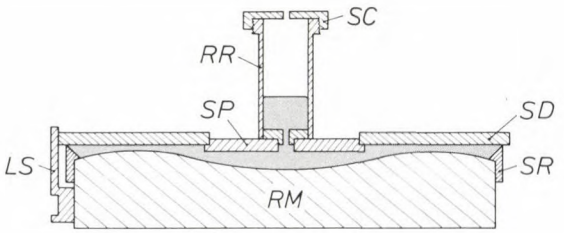


Fig. 5. Construction of the master mould. *RM* profiled base plate (glass). *SD* substrate plate (glass). *SR* sealing ring (silicone rubber). *LS* location stop. *SP* flexible sealing plug (polythene). *RR* resin reservoir (glass). *SC* metal screw cap with air hole.

Table I. Some mechanical properties of the cured epoxy-resin material used for the corrector plates.

Specific gravity	1.2
Compressive strength	70 MN/m²
Tensile strength	40 MN/m²
Izod impact (notched specimen)	4 MN/mm
Shore hardness (D scale)	75
Flexural strength	140 MN/m²
Water absorption	0.2 vol% in 24 hr
Linear expansion coefficient	$6\text{-}10\times10^{-5}$/°C

relating to viscosity, curing temperature, curing shrinkage, rigidity, refractive index, softening point and coefficient of expansion. The spectral transmission and the accuracy of reproduction of small details are more than adequate for the application.

Since the two epoxy resins have different refractive indices (1.582 and 1.574 at 589.6 nm) a tailor-made refractive index can be obtained by changing the mixing ratio. The refractive index of the mixture finally used is 1.576 at 589.6 nm. The difference in refractive index for light at 656.3 nm and 486.1 nm is only 0.016, so that the chromatic aberration of the plate is low. *Table I* shows the mechanical properties of a cured resin layer of the composition used.

The glass substrate has two functions. It acts as a support for the resin layer, and it reduces the amount of X-radiation originating from the projection tube to an acceptable level.

The glass substrate must be carefully prepared to obtain optimum adhesion between it and the resin layer. The standard procedure finally followed for cleaning the glass surface consists of a sequence of washing, rinsing, wiping and drying; agents used in these operations were acetone, Teepol, a chalk suspension, demineralized water and isopropyl alcohol. After the cleaning, a bond promoter is applied to the glass substrate to improve the long-term stability of the adhesive bond.

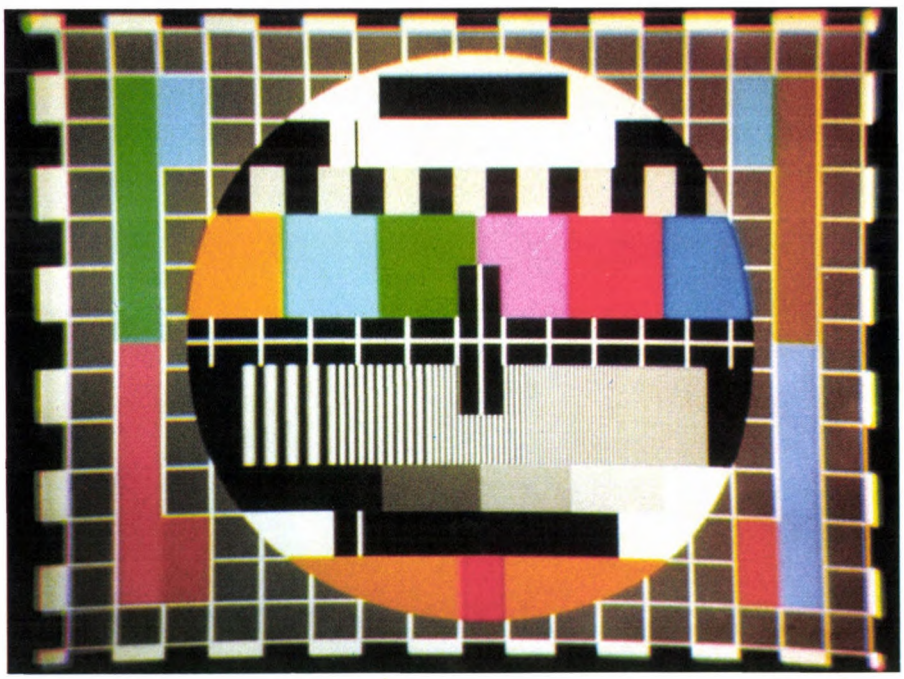


Fig. 6. Projection of a standard test pattern obtained with an experimental projector fitted with the new type of correction plate. The cathode-ray tubes, manufactured in pilot production, have a screen of diameter 75 mm. The distance between projector and screen was 2 m, the picture diagonal 1.25 m. The maximum resolution demonstrated corresponds to the full bandwidth of the television input signal (4.8 MHz).

While the bond between the resin layer and the substrate must be as good as possible, the resin layer, on the other hand, must adhere to the mould as little as possible. For this reason a release layer is applied to the master mould after it has also been subjected to a special cleaning process. A number of commercially available types of silicone wax dissolved in a volatile solvent were found to be suitable for this layer. The thickness of such a layer is so very small (< 10 nm) that the surface quality of the glass master mould is accurately reproduced in the surface of the resin layer.

Fig. 5 shows the construction of a mould that is found to be highly suitable for casting the correction plates. The profiled section of the mould — which is made of glass but can in principle be made from other materials, such as aluminium — acts as the base; the glass substrate forms the top of the mould cavity. The central hole in the substrate is sealed by a flexible polythene disc through which an enclosed cavity is connected to the mould cavity. This reservoir contains 100 grams of resin mixture.

The master mould is gradually filled by the resin flowing in under its own weight, while the resin drives out the air in the mould through a silicone rubber ring. This ring has a tapered edge so that it acts as a non-return valve and the edge of the circumference of the resin layer to be formed will not be too steep. After pouring, the resin should be cured for a minimum of 4 hours at 30 °C.

If desired, the process can be accelerated by injecting the resin into the mould under pressure. The curing process can also be shortened by using a larger quantity of curing material or by selecting a higher cure temperature. A process is being considered that will reduce the cure time to 10 minutes.

To remove the correction plate from the master mould, the rubber sealing ring is taken out and a screw with vertical movement is clamped to the circumference by a clamping strip. When the screw is turned it pushes against the sloping resin surface, producing a localized tensile force on the release layer at this edge. At a force of some tens of newtons the resin layer begins to separate; this is confirmed by the appearance of Newton's rings. The complete correction plate can then be released without difficulty.

The final operation consists in removing the polythene plug and the central part of the resin layer.

It should be borne in mind that there is always some shrinkage of the synthetic-resin material. The linear contraction is 2%, with 60% of this occurring in the liquid phase. There is compensation for this loss of volume by a supply of resin material from the part below the polythene plug. The residual contraction of 0.8% results in a proportionate reduction in the depth of the final profile. To compensate for this the depth is overdimensioned when the master mould is made.

Fig. 6 gives an impression of the quality of a picture projected on a screen of about 1 m^2 with the new type of correction plate.

Medical electroradiography — its potential and limitations

K. R. Peschmann

'Electrophotography', which is now well known because of its widespread application in copying machines, is also employed in medical diagnostics, where it is called 'electroradiography'. One important reason for research on electroradiography is that it makes no use of silver, unlike conventional X-ray photography. Another attraction is the range of possibilities offered by the developing process: the picture information can be processed to improve the 'readability'.

The studies in electroradiography at the Philips research laboratories in Aachen have been mainly concerned with the possibilities and the limitations both in the formation of the latent charge pattern — sensitivity and resolution — and in making this visible (the 'development'). The article deals mainly with the system in which the X-ray radiation is detected with the aid of a compressed gas.

Introduction

Electrophotography has become generally known as a result of the widespread use of copying machines. The subdivision of electrophotography in which an X-ray field is used to provide the input information is known as electro*radiography*. Because electrophotography does not use silver to produce pictures, this in itself is sufficient reason for investigating the possibility of using it for medical diagnostics. Even though the combination of X-ray image intensifiers and television systems and the associated magnetic-memory technique has been with us since about 1960, the amount of silver-based film used for X-ray photography in North America in 1974, for example, still averaged out at 0.4 m^2 per head of the population. It should, of course, be pointed out that recovery of silver from the used silver bath is relatively simple and as long as the price of silver keeps on increasing recovery will be used more and more. Silver can also be recovered from old X-ray photographs and this is being done, particularly in the United States.

In medical radiography internal structures in the human body are made visible by using X-rays and the principle of shadow projection. In electroradiography the X-ray radiation attenuated by the object is absorbed by an X-ray photoconductor, thus creating charge carriers. Under the influence of an electric field the charge carriers come to rest on a dielectric film (called an insulating 'foil') where they produce an (invisible) charge pattern of the tissue structure being photographed. This pattern is made visible by means of 'electrophotographic development'.

Another important feature resulting from the changeover from conventional to electrophotographic image recording is that the electrophotographic developing process can be varied and this presents us with an extra 'dimension'. It offers us a relatively simple method of 'picture processing'. It is possible, for example, to attenuate the background caused by scattered radiation. Another well-known example of the processing of picture information is contour enhancement in the reproduction of weak contrasts in xerography. This is accompanied by a loss of information, however, when relatively large homogeneous parts and areas of average-to-high contrasts in the object are being reproduced. In certain cases the nature of the picture may also be altered to such an extent that considerable experience is necessary for interpreting it.

Whatever the method used, the patient should not be exposed to a higher dose of radiation. The objective of our investigation is therefore to develop systems in which the dose of radiation required is not greater than for film radiography with an intensifying screen, but in fact rather smaller. Xerography is already being used for mammography, particularly in the U.S.A. This system is somewhat insensitive,

Dr K. R. Peschmann was formerly with Philips GmbH Forschungslaboratorium Aachen, Aachen, West Germany.

however, so that the required dose of radiation is high. Experiments with gas ionography — an electroradiographic system with a gas as the photoconducting medium — have yielded results that are more promising; it has been shown that the radiation dose here can be reduced compared with conventional systems.

The objective of the investigation is not so easy to define for the 'picture quality' in general as it is for the required radiation dose. Many factors affect the picture quality; specially adapted systems have to be used for the various methods of investigation. In practical terms, this means that the basic system of X-ray source and recording medium must be optimized for all applications.

Any of the possible recording systems is subject to certain limitations. There are two main reasons for this. Firstly, the radiation is not completely absorbed, i.e. not all the X-ray photons contribute to the formation of the image. More primary photons are therefore necessary than in an ideal system. Secondly, the processes used for image formation are subject to noise, which degrades the resolution of details.

A radiographic system is said to be *limited by quantum noise* if the resolution of detail in the image is limited by statistical fluctuations in the flow of X-ray photons and not by the sources of noise inherent in the system or by effects that reduce the sharpness of the image. A calculation we made for mammography [1] shows that about 10^8 photons per cm^2 must be incident on the recording medium for microcalcifications with a cross-section of 150 μm to be detected. This is equivalent to a dose of 7.5 mR (1 R = 2.58×10^{-4} coulombs/kg) for a mean quantum energy of 26 keV. If it is assumed that the object has a mean transmission of 6% then the dose of radiation to which the patient is exposed is about 120 mR. The calculation holds for a system limited by quantum noise in which the detective quantum efficiency is set at 1 (we shall return to the concept of detective quantum efficiency later) and it has been assumed that the system operates linearly, i.e. the optical density is proportional to the flux density of the X-ray photons. It is easy to see that the required dose becomes greater as the particles to be imaged become smaller.

Table I. Comparison between three electroradiographic systems

Production of an electroradiograph	High-pressure gas-ionography system	Lead-oxide/gas-layer system	Xerographic system
X-ray photon → absorption	absorbing medium: 1 cm layer of xenon, krypton or bromine-Freon (10 bars)	absorbing medium: lead-oxide crystals in a binder layer, effective thickness 80 μm	absorbing medium: 135 μm amorphous selenium layer evaporated on aluminium; surface charge applied by a corona discharge
high-energy electrons and photons → thermalisation	deceleration, giving rise to 'clouds' of positive ions and low-energy electrons	deceleration, giving rise to mobile electrons and holes	
charge carriers → drift and transfer	drift of (+) ions and (−) electrons or (−) ions in the electrical field and 2-dimensional distributed charging of a foil	drift of electrons and holes in the field; formation of 2-dimensional distributed microscopic gas discharges and charging of transparent foil	drift of charge carriers and 2-dimensional compensation of the corona charges
electrostatic image → electrophotographic development	electrophoretic precipitation of charged pigment particles, e.g. suspended in an apolar liquid		making the contours of the charge pattern visible by means of charged pigment powder; transfer of the pigment image to paper
electroradiograph	similar to a radiograph, transparent image		xeroradiograph

In the sections that follow we shall look first at the physical processes that give rise to the formation of an electroradiograph, and then at two systems of electroradiography, the high-pressure gas ionographic system [2] and the lead-oxide/gas-layer system, for which we have made a special study of the sensitivity and resolution. The possibilities offered by the

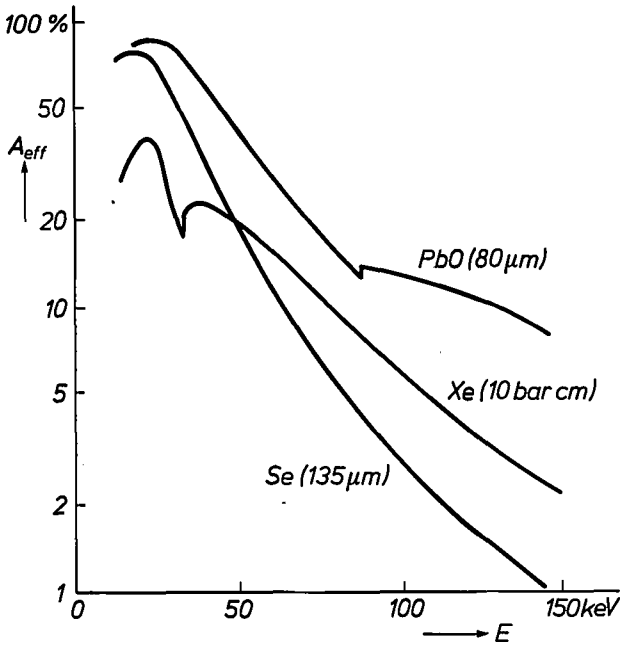


Fig. 1. Calculated effective X-ray absorption A_{eff} as a function of the photon energy E in photoconductors used in different radiographic systems. The fact that part of the fluorescent radiation and the scattered radiation do not contribute to the image information is taken into account. The dip on the left can be attributed to the absorption of the electrode at the entry side. A striking feature is the relatively low effective absorption of the compressed xenon gas (maximum about 40%). The 80 μm PbO layer has the highest absorption.

development of the charge pattern will then be examined; a great deal of the knowledge obtained in this field is based on investigations carried out in our laboratories. Finally, a more detailed examination is made of the advantages and disadvantages of electroradiographic systems.

Electroradiographic systems

In this section an explanation of the operation of electroradiographic systems will be given, with the high-pressure ionographic system and the lead-oxide/gas-layer system as examples. The two methods are compared in *Table I*, which also includes selenium xeroradiography for completeness. This system will not be discussed further here, however [3].

The left-hand column of the Table indicates the stages in the process of making an electroradiograph. The other columns give special features of the three systems. These columns start with a short description of the photoconductor, which absorbs the X-ray radiation that passes through the object under examination. In ionography the photoconductor consists of a compressed heavy inert gas, e.g. xenon or krypton, while in the PbO system the photoconductor is in the form of a layer of lead-oxide crystals in a binder that has an effective thickness of 80 μm. As a result of the absorption processes a spectrum of high-energy photons and electrons is produced. The electrons — together with the L fluorescence photons from the lead oxide and the L and K fluorescence photons from the selenium — are slowed down by the absorbing medium, each high-energy electron generating several hundred to two thousand secondary charge carriers. The calculated effective absorption has been plotted in *fig. 1* for the three systems. As the figure shows, the PbO system with the 80 μm layer has the highest effective absorption in the total energy range, up to almost 90%, whereas only 40% of the radiation is used effectively in the high-pressure gas system with xenon. (Hasty conclusions should not be drawn from this figure, since the effective absorption is only one of the criteria used to assess the electroradiographic systems.)

An electric field pulls the mobile secondary charges in the direction of the electrodes. In the high-pressure gas system the charge transport ends when the charge carriers have reached an insulating foil. In the PbO system this transport is terminated as soon as the charge carriers arrive at the surface of the photoconductor, provided they have not previously been trapped. Charge transfer takes place from the surface of the photoconductor through the gas layer to an image-bearing foil. In both cases a charge pattern has been formed on a foil — an electrostatic surface-charge distribution whose density is proportional to the X-ray radiation field attenuated by the object. The final stage is the conversion of the charge pattern into a visible picture.

Ionography with a compressed gas

Secondary spectrum and detective quantum efficiency

Fig. 2 is a diagram showing how a charge pattern is formed by means of ions of a compressed gas. The X-ray photons to be detected enter the gas space G, which contains a gas such as xenon as the absorbing medium, through the pressure-tight window electrode W. The secondary radiation generated by the internal photoelectric effect and by scatter consists of high-

[1] K. R. Peschmann, Empfindlichkeitsgrenze elektroradiographischer Systeme bei Mammographiestrahlung, Tagungsband Medizinische Physik, Hüthig Verlag, Stuttgart 1978.
[2] H. E. Johns, A. Fenster, D. Plewes, J. W. Boag and P. N. Jeffery, Gas ionization methods of electrostatic image formation in radiography, Brit. J. Radiol. 47, 519-529, 1974.
[3] More information on selenium xeroradiography is given in: J. W. Boag, Xeroradiography, Phys. Med. Biol. 18, 3-37, 1973;
F. V. Allan and K. H. Reiss, Xeroradiography and ionography, Med. Progr. Technol. 3, 143-148, 1976;
F. W. Schmidlin, V. Dryden, M. Canon and F. Yang, Physical theory of xeroradiography and the observation of photon noise, Opt. Engng. 15, 344-348, 1976.

energy electrons and photons whose calculated spectra are reproduced in *fig. 3*. The area of the regions bounded by these curves is proportional to the energy of the particular secondary-radiation component. The figure clearly shows that the absorbed radiation energy comes mainly from the fast electrons. When these are decelerated in the gas a number of ion/electron pairs are created, and these are charge carriers that will form the charge pattern. The photons (dashed curves) that are simultaneously formed with the fast secondary electrons are lost because they cover a few millimetres in the gas and are therefore not absorbed until they are a considerable distance away from the point of origin — and may even reach the wall of the ionographic chamber.

The extent to which the radiation incident on the chamber is used for the information present in the X-ray picture is expressed by the detective quantum efficiency DQE. The value of this indicates directly the radiation dose required in an actual system as compared with that required in an ideal system (DQE = 1). If for example DQE = 0.25, then the dose of radiation must be four times as high if the same image information (limited by quantum noise) is to be obtained as in an ideal system. The detective quantum efficiency is defined as follows [4]:

$$\mathrm{DQE}(E) = \left[\frac{n_0(E)}{s_0(E)}\right]^2 \Big/ \left[\frac{n_{\mathrm{det}}(E)}{s_{\mathrm{det}}(E)}\right]^2, \quad (1)$$

in other words, as the square of the noise-to-signal ratio of the X-ray radiation incident on the chamber, divided by the square of the noise-to-signal ratio of the signal produced by the charges generated in this chamber. *Fig. 4* shows the result of a calculation of this quantity for xenon or krypton as the medium [5]. The curves show that under the conditions as described in the caption to the figure, a maximum DQE of 40% is obtained with a photon energy of 21 keV and that for xenon there is a second maximum of 23% at 50 keV, whereas the DQE for krypton is only half as high. For mammography, which employs photon energies of less than 40 keV, krypton is more suitable as the medium than the much more expensive xenon. For the radiography of other organs harder radiation is used. In these cases xenon is therefore more suitable, and as we shall see, it has a greater sensitivity.

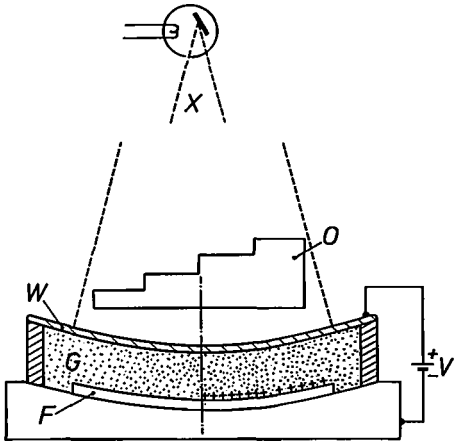


Fig. 2. Diagram of the high-pressure gas ionographic system. *X* X-ray radiation. The X-ray radiation attenuated by the object *O* enters the ionography chamber through the window electrode *W* and is absorbed in the gas space *G*. The charge carriers released during the absorption are pulled by the electric field towards the insulating foil *F*, forming a charge pattern on it. The left-hand and right-hand halves of *F* represent the situation before and after exposure. The spherical shape of the chamber ensures that the paths of the X-ray photons and the direction of the electrical field are parallel, so as to prevent any adverse effects on the image sharpness.

A word or two of explanation should be given about the derivation of the somewhat complicated-looking equation (1)[6]. The average number of X-ray photons in a small energy interval ΔE around E incident on one square centimetre of the ionography chamber during an exposure is represented by $\Phi(E)\Delta E$. For the associated energy signal $s_0(E)$ we have:

$$s_0(E) = \Phi_0(E)\,\Delta E E.$$

Because the photon flux is subject to fluctuations the standard deviation of the number of X-ray photons is $[\Phi_0(E)\,\Delta E]^{\frac{1}{2}}$ for Poisson statistics. This gives:

$$n_0(E) = [\Phi_0(E)\,\Delta E]^{\frac{1}{2}} E,$$

for the standard deviation of the energy signal. The ratio n_0/s_0 is called the natural noise-to-signal ratio of the X-ray radiation detected.

The calculation of the noise-to-signal ratio $n_{\mathrm{det}}(E)/s_{\mathrm{det}}(E)$ for the charges generated in the chamber is a little more difficult. Different absorption cross-sections have to be taken into account because they determine the probability of the individual absorption processes. Because the radiation is attenuated by the window, incomplete absorption in the gas, scattering processes and fluorescence radiation (see fig. 3) the associated energy signal is weaker than the signal that would be supplied directly by the X-ray radiation detected. Because the noise-to-signal ratio has also become larger, DQE always has a value smaller than one.

Sensitivity

While the DQE is a criterion for assessing the effectiveness of X-ray radiation in the image formation, the sensitivity of the ionographic chamber is a criterion for assessing the density of the surface charge produced on the foil. A high sensitivity means a high charge density, which in turn is a precondition for high optical density. A high maximum value for the optical density or a wide range for the density scale makes it possible to obtain developed images that contain a large number of grey levels.

From the definition of the radiation (ion) dose expressed in röntgens (R), it can be shown that N_1

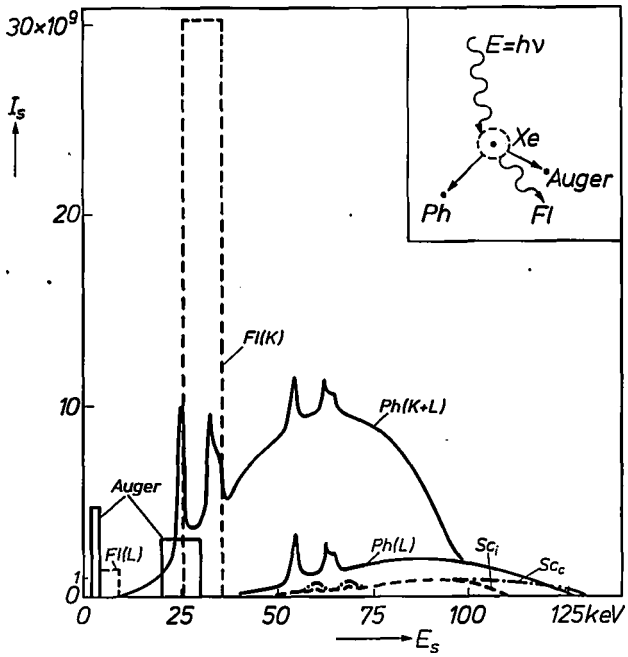


Fig. 3. Calculated energy spectra of the secondary electrons and photons in xenon. When an X-ray photon of energy E collides with a xenon atom *Xe* the internal photoelectric effect produces secondary radiation that consists of photoelectrons *Ph*, Auger electrons *Auger* and fluorescence photons *Fl* (*inset*). For the kinetic energy of the photoelectrons $E_{kin} = E - E_{binding}$.

In the graph, the intensity I_s (number of quanta per energy interval × quantum energy) of the various secondary radiation components (in arbitrary units) is plotted against the energy E_s so that the areas of the regions bounded by the curves are proportional to the energy of the associated components. Only the photoelectrons and the Auger electrons are effective; the components indicated by the dashed curves do not contribute to the image information. *Ph*(K), *Ph*(L) photons from K and L shells. *Fl*(K), *Fl*(L) K and L fluorescence photons. Sc_c and Sc_i scattered coherent and incoherent photons.

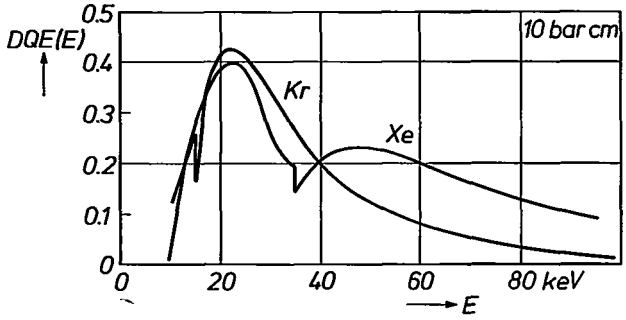


Fig. 4. Calculated detective quantum efficiency DQE(E), equation (1), as a function of the photon energy E for xenon and krypton of 10 bar cm. Graphite-fibre-reinforced epoxy resin is used as the window for the incident radiation (thickness 8 mm). The maximum DQE is about 40% for the two gases; at 50 keV xenon has a second maximum of about 23%. For mammography in which soft radiation is used krypton is preferred because it is cheaper than xenon; with the harder radiation used for example for chest X-rays xenon is more suitable.

photons with an energy E (in keV) are equivalent to a radiation dose of

$$1.425\times10^{-11}\,\mu_{tot,a,L}(E)\,N_1(E)\,E \text{ röntgens}, \quad (2)$$

where $\mu_{tot,a,L}$ represents the linear photoabsorption coefficient of air while the indices 'tot' and 'a' relate to the method used for determining μ. These N_1 photons generate a quantity of charge (either positive or negative) in the gas of the order of

$$N_1(E)\,E\,[1 - A_F(E)]\,f(E)\,A(E)\,\frac{1}{W}\,e \text{ coulombs}, \quad (3)$$

where $A_F(E)$ is the absorption of radiation in the input window of the ionographic chamber, $f(E)\,A(E)$ is the absorption in the gas, reduced by the quantity of fluorescent radiation, and W is the conversion value for the medium. This value indicates the quantity of radiation energy (in eV) necessary for the generation of a single pair of charge carriers. For xenon W is equal to 22 eV/pair, which means, for example, that a photoelectron with an energy of 30 keV creates a maximum of 1400 pairs of charge carriers. In equation (3) e represents the charge of the electron ($e = 1.6\times10^{-19}$ coulomb). The sensitivity $S(E)$, expressed as the quantity of charge released per milliröntgen, is equal to the ratio of (3) and (2):

$$S(E) = 5.10\times10^{-13}\,\frac{[1 - A_F(E)]\,f(E)\,A(E)}{\mu_{tot,a,L}(E)}\,\frac{C}{mR}. \quad (4)$$

The conversion value W from the expression (3) is included in the constant; the coefficient $\mu_{tot,a,L}$ for air can be calculated from a table [7]. The calculated curve of $S(E)$ for an absorbing layer of xenon or krypton is illustrated in *fig. 5a*; it has been assumed that the absorption in the window is negligible.

The sensitivity S to polychromatic radiation with a distribution $N_1(E)$ is obtained by integrating (4) with respect to the energy. Fig. *5b* shows the variation for diagnostic X-ray radiation when xenon and krypton are used. The lower curves were obtained with the radiation from an industrial X-ray tube with a tungsten anode — the same kind of radiation as is used in mammography.

The sensitivity of the high-pressure gas-ionographic system can be increased considerably with the aid of electron multiplication by collision, by employing the Penning effect [8]. In this case a dose of 10 μR is sufficient to produce a charge density of 2×10^{-9} C/cm². In comparison with fig. *5b* this amounts to an improvement by a factor of 100. This method, however, has the disadvantage that the resolution is reduced to 1 to 2 line pairs/mm because of the lateral diffusion of the electrons (see also fig. *9b*).

[4] A similar definition of detective quantum efficiency is found in conventional photography; see: J. C. Dainty and R. Shaw, Image science, Academic Press, London 1974.

[5] K. R. Peschmann, Xenon ionization X-ray detectors, to appear in: T. H. Newton (ed.), Radiology of the skull and brain, part 5 (section II), Mosby, St. Louis 1979/80.

[6] A detailed derivation of the expression for detective quantum efficiency is given in [5].

[7] E. Storm and H. I. Israel, Photon cross sections, in Nuclear data tables A7, 565-681, 1970.

[8] K. R. Peschmann and G. Grosche, Amplification and entailed resolution degradation in high-pressure gas ionography, Med. Phys. 4, 202-207, 1977.

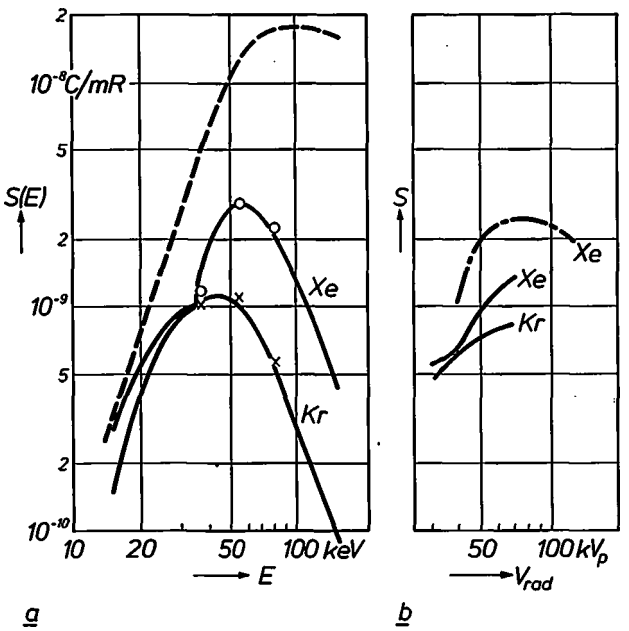


Fig. 5. Sensitivity of an ionography chamber 1 cm thick filled with krypton or xenon (10 bars). Voltage across the ionography chamber 12 kV. *a*) Spectral curve for the sensitivity $S(E)$. In the calculation of the curves the K and the L fluorescent radiation is considered to have been lost and the absorption in the window has been neglected. The test points were obtained with highly filtered radiation with photon energies of about 37 keV, 55 keV and 80 keV and a beryllium window whose absorption was only a few per cent. For comparison, the dashed-line curve gives the calculated sensitivity of an 'ideal' chamber (complete absorption in the gas and no loss in the window; the numerator in equation (4) is then 1). The shape of this curve is determined by the definition of the ion dose (röntgen). *b*) The sensitivity S as a function of the tube voltage V_{rad} for diagnostic X-ray radiation. The solid curves were obtained with the radiation from an industrial X-ray tube with a tungsten electrode and its own filter, i.e. with radiation as used for mammography. The phantom filter was a 4 cm thick layer of acrylic resin. The chain-dotted curve was obtained with radiation from a tube with a tungsten electrode operating at 10 mA with a 6-pulse generator; the radiation was filtered by 1 mm of aluminium and 15 cm of water to simulate the human body.

Resolution

The resolution that can be obtained in gas ionography is always limited by two physical effects. The first effect has been illustrated in *fig. 6*. When a photoelectron is decelerated, a 'cloud' of pairs of secondary charge carriers is formed and the size of this cloud depends on both the initial energy of the electron and the nature and pressure of the gas. It has been estimated that the distance between the point at which the primary photon is absorbed and the centre of gravity of the ion cloud produced in xenon at a pressure of 10 bars by an electron with an energy of 30 keV is about 0.15 mm. It is shown diagrammatically how a jump in the intensity profile results in a continuous change of the density in the charge pattern.

The second effect is the lateral diffusion of the electrons and ions during their drift; see *fig. 7a*. The electrons diffuse farther than the ions, since for each elastic collision with a xenon atom they can only lose one millionth part of the energy they have received from the applied field. This leads to a very fast, statistically distributed movement (with an electron temperature of several tens of thousands of degrees kelvin), in which the distance covered is 300 times as long as the path in the direction of the field. Because of this, a charge pattern consisting of electrons results in a radiographic image that is much less sharp than the ion image. It is possible to improve the situation, however, by adding to the rare gas a small quantity of another gas that has the property of retaining an electron at a collision. A negative ion is then formed whose lateral drift behaviour is much the same as that of a positive xenon ion; see fig. 7*b*. A simple model has shown that the blurring of the ion-formed charge pattern caused by the diffusion is independent of the gas pressure and inversely proportional to the square root of the applied field.

To determine how the resolution of an ionographic system depends on the electric field we have prepared radiographs of a line structure made of lead foil

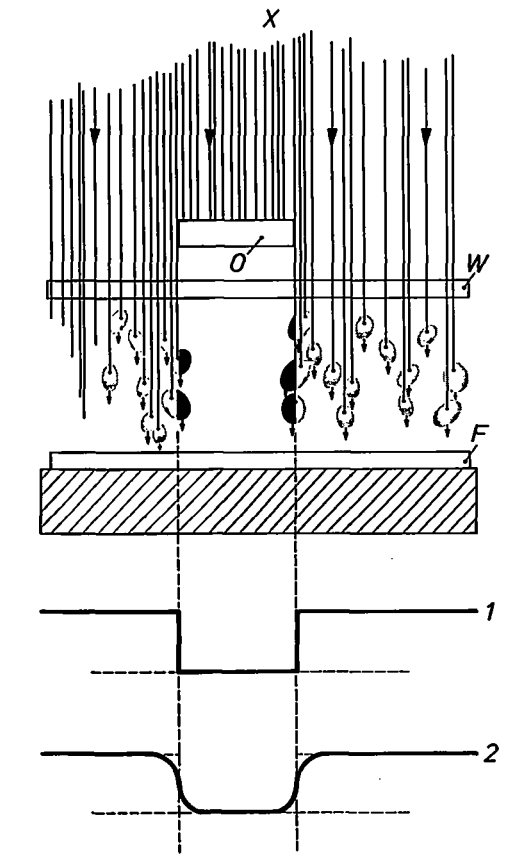


Fig. 6. Limitation of the resolution resulting from the radius of action of the electrons (schematic). When a fast photoelectron is slowed down, a 'cloud' of pairs of secondary charge carriers is formed. If this takes place in the vicinity of the geometrical edge of the shadow, the 'cloud' stretches out over this edge. In this way an abrupt change in the intensity in the radiation field (curve *1*) becomes a continuous changeover in the surface charge density of the latent image (curve *2*). *X, O, W, F* as in fig. 2.

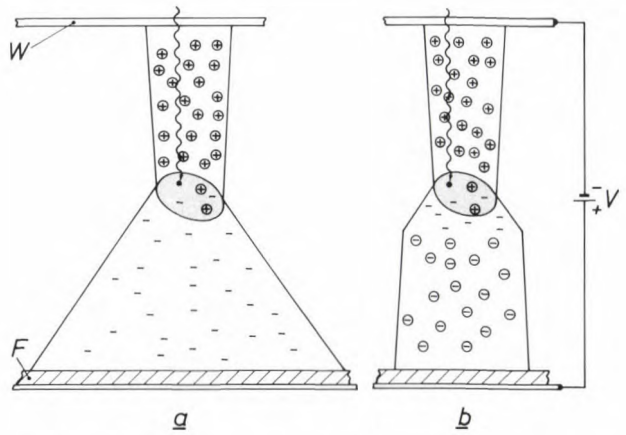


Fig. 7. Limitation of the resolution by lateral diffusion of charge carriers (schematic). The charge carriers that are formed when the fast electrons are decelerated move under the influence of the voltage *V*. In this case the positive ions go to the top and the negatively charged particles to the bottom. *W*, *F* are the same as in fig. 2. *a*) The electrons (−) diffuse further than the ions (⊕), since for each collision with a xenon atom they can only lose a small fraction of their kinetic energy. This has an adverse effect on the resolution. *b*) This disadvantage can be overcome by adding to the working gas a small quantity of a gas that retains the electrons at a collision. The negative ions (⊖) formed in this way have much the same kind of lateral movement as the positive xenon ions.

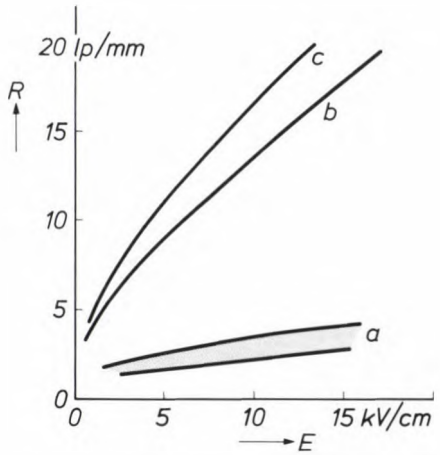


Fig. 8. Experimentally determined resolution *R* (in pairs of lines per mm) of an ionography chamber (thickness of gas layer 1 cm, working gas xenon (7-8 bars), no gas amplification by electron multiplication), as a function of the electric field-strength *E*. Primary radiation: 35 kV_p, tungsten anode; object of measurement: line structures of 25 μm and 50 μm lead foil ('*Funk-raster*'); fine-grained liquid developer. Region *a*: electron image; curve *b*: positive ion image; curve *c*: negative ion image, formed by the addition of an electron-absorbing gas.

25 μm thick ('*Funk-Raster*'). A magnifying glass or a low-power microscope was used to determine the resolution of ionographs prepared with a linear very fine-grained liquid developer. The result is illustrated in *fig. 8*; the experimental conditions are given in the caption. Reproductions of subsequently enlarged ionographs are shown in *fig. 9*. It is possible to distinguish more than eight pairs of lines per mm in the ion image (fig. 9*a*) and one pair of lines per mm in the electron image (fig. 9*b*).

[9] M. Pfeiler and G. Theil, Die Übertragungsfunktion in der Radiologie — ein kleines Bildkompendium, Röntgenpraxis **28**, 205-218, 1975.

The modulation transfer function [9] is a useful criterion for assessing a radiological system. It gives a quantitative description of the extent to which an image system is capable of reproducing certain details of the object. The modulation transfer function of our experimental ionographic system is illustrated in *fig. 10*.

The lead-oxide/gas-layer system

Sensitivity

In the section on 'Electroradiographic systems' the PbO system has already been mentioned briefly in Table I; *fig. 11* is a diagram of this system. Electrode

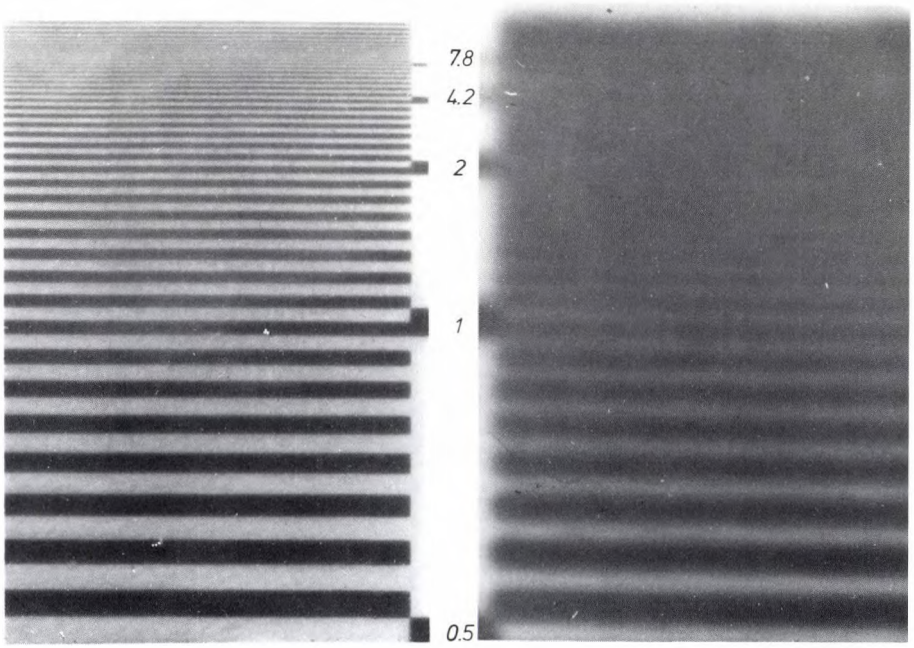


Fig. 9. Enlarged reproductions of ionographs (radiation 45 kV_p) of a 25 μm '*Funk-raster*' made visible by a linear and relatively insensitive liquid developer. The numbers give the density of the lines in pairs of lines per millimetre. *a*) Charge pattern of negative ions, dose 10 mR. *b*) Charge pattern of electrons, dose 0.1 mR. This picture is much less sharp because it is a charge pattern of electrons, and use is also made of the Penning effect (gas amplification 110×) [8]. This has the advantage that the sensitivity is greatly increased (dose 0.1 mR as compared with 10 mR in (*a*)).

E, foil F and gas layer G are almost completely transparent to the radiation detected. The absorption takes place in PbO crystals, which are held together by a binder (layer L). In fig. 1 we saw that this 80 μm layer has a high absorption efficiency. This satisfies an important condition if a layer of this kind is to be used as a photoconductor in a radiographic system, because the sensitivity of a system depends in part on the absorption.

Another precondition for high sensitivity is to have as low a conversion energy as possible for generating image-forming charge carriers. In this case, i.e. in a solid, this means that the loss due to the disappearance of charge carriers by capture in traps and by recombination has to be kept to a minimum. Experiments [10] on single crystals of different semiconductor materials have shown that this conversion energy is about three times the energy gap. In the case of (crystalline) tetragonal PbO this rule of thumb gives us a value of about 7 eV. In practice, however, a higher value will be found for several reasons (capture of charge carriers, recombination, non-single-crystal nature of the layer, charge transfer via gas layer).

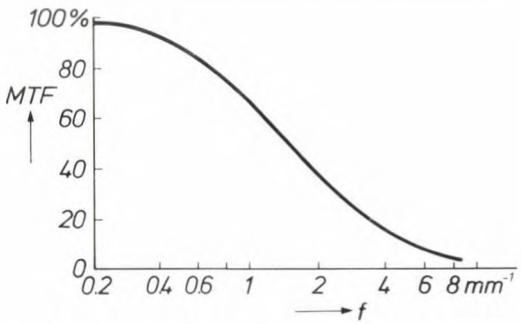


Fig. 10. Modulation transfer function MTF of the experimental ionography system as a function of the spatial frequency for linear development. Working gas xenon (8 bars); thickness of gas layer 1 cm; voltage 16 kV; X-ray radiation of 40 kV_p; short focus, focal length 2 m. The calculation was made by Fourier transformation of the X-ray image of a 75 μm gold wire [*].

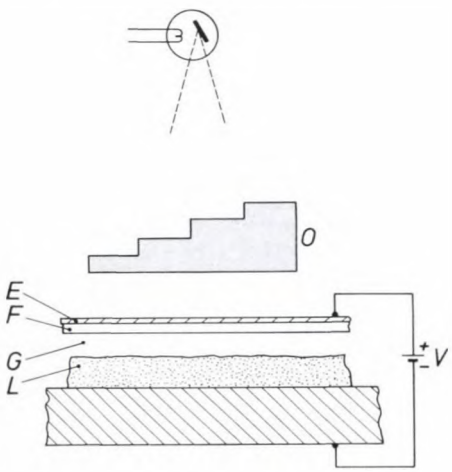


Fig. 11. Diagram of the PbO/gas-layer system. The electrode E, the foil F and the gas layer G are transparent to the X-ray radiation. The X-ray radiation attenuated by object O is absorbed in the PbO binder layer L. The mobile secondary charge carriers formed move under the influence of the electric field produced by applying a high voltage V between the electrodes. Charge transfer takes place from the surface of the photoconductor to the foil, producing a charge pattern on the foil.

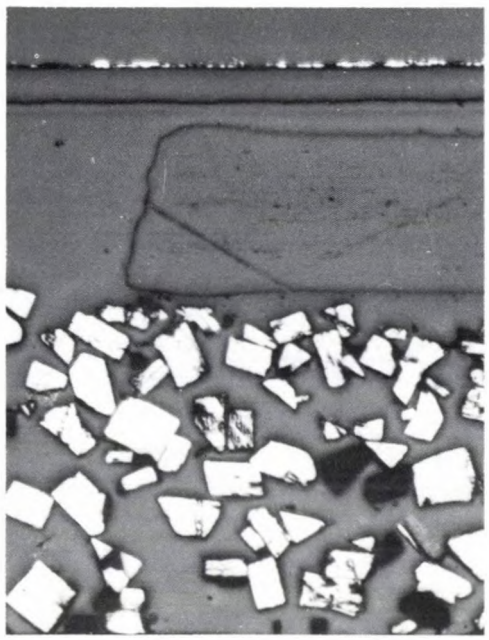


Fig. 12. Micrograph of cross-section of the binder layer. The PbO crystallites are 10 to 20 μm in size; the layer also contains about 1% of organic binder and about 70% of air. Above the binder layer is the gas layer, about 100 μm thick, which is bounded by the foil. The metal substrate cannot be seen. To prepare the polished section the gap for the gas layer was filled with an epoxy resin.

In the laboratories in Aachen the X-ray photoconductor layer was prepared by sedimentation on a metal substrate of very pure tetragonal lead oxide with a mean crystallite size of 10 to 20 μm from a resin solution [11]. This method produces layers that contain, in addition to lead oxide, about 1 wt% of organic binder and about 70 vol% of air. *Fig. 12* illustrates a cross-section of such a photoconducting layer. Because of the sponge-like structure it is impossible for the charge pattern on the photoconducting layer to be electrophotographically developed directly. This is why the system shown in fig. 11 [12] has been used. At a distance of about 100 μm from the layer is an insulating foil 6 to 30 μm thick that captures the charge pattern. At the back of the foil is an electrode with a relatively high sheet resistance (about 10^7 Ω/□). If a voltage of 1750 to 2000 V is applied between this electrode and the photoconductor substrate, the X-ray exposure causes the charge to be transferred to the foil, thus producing the image. For the present no quantitative model has as yet been devised for the physical processes that occur, but the pre-ionization of the gas (usually air at normal pressure) by secondary radiation from the

lead oxide could be an important factor. Under the conditions described here an effective conversion energy with a reproducible value of about 20 eV has been measured.

Resolution

Like the ionography system, the PbO system is also subject to unavoidable sources of noise that have an adverse effect on picture quality. One of the reasons for the noise is the fact that on average there are 5 to 6 crystallite layers one above the other in the binder layer. As a result of the exponential attenuation of the radiation penetrating the binder these layers make very different contributions to the formation of the charge pattern. Other reasons [13] are that the crystallites are not uniformly distributed in the binder ('agglomeration') and the amount of charge generated in each of the crystallites differs. Finally, the mechanism of the transfer of this charge to the foil via the gas layer is also responsible for a considerable amount of noise. The noise in the lead-oxide system is the factor that makes the detective quantum efficiency much lower than the quantum absorption shown in fig. 1. The picture in *fig. 13* is a radiograph of a lead-foil *'Funk-Raster'* obtained using a lead-oxide system with a dose of 20 mR. The limit of resolution of detail is equivalent to a resolution of 10 line pairs per mm.

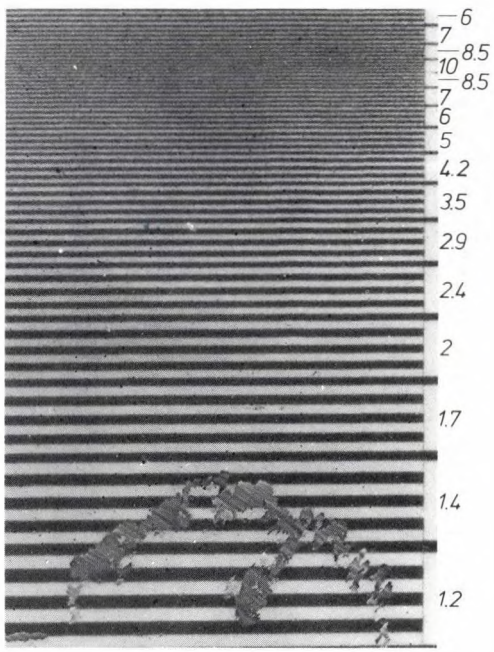


Fig. 13. Reproduction of a radiograph of a 'grid' of lead foil as obtained with a lead-oxide system. The object was irradiated with a dose of 20 mR. It is just possible to distinguish 10 pairs of lines per mm. The grainy structure (image noise) is caused by imperfections in the lead-oxide layer and the charge-transfer process.

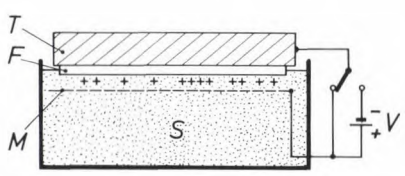


Fig. 14. Diagrammatic cross-section of an apparatus for liquid development. *T* is the electrode with foil *F*, which carries the charge patterns. The developer electrode *M* ensures that as many lines of force as possible extend vertically from the charge pattern into the developer suspension *S*. The particles in the suspension move along the lines of force and attempt to neutralize the charge pattern. By applying a bias voltage to *M* it is possible to suppress any background caused by scattered radiation (fig. 15).

The development of grey levels

Principle

The last stage in the production of the electroradiograph is making the latent charge pattern that has been produced on the foil visible by electrophotographic development. The method of development has a marked effect on the total sensitivity of the electroradiographic system and on the image quality. To clarify the concept of 'total sensitivity of the system' it should be noted that, in addition to the 'sensivity' that has been discussed in the previous section and is expressed in coulombs/milliröntgen, a 'development sensitivity' also has to be taken into account: this refers to the deposition of pigments per coulomb.

Besides the well-known developing method that uses a cloud of powder, there is also another electrophotographic developing procedure, 'liquid development', which we have investigated. In an apparatus such as the one shown in *fig. 14* the charge pattern on the foil *F* is brought into contact with the suspension *S* of electrically charged particles of pigment in an insulating liquid. The liquid also contains ions, together with additives that have a complicated composition for regulating the charge of the pigment particles and maintaining stability [14]. During the development process the charged particles move along the lines of force of the electric field, which originates from the charge pattern, and settle on the insulating foil. The electrode *M*, the 'developer electrode', which is a short distance from the charge pattern in

[*] The calculation has been carried out by Ir F. Timmer, Philips Medical Systems Division, Best, the Netherlands.

[10] C. A. Klein, Bandgap dependence and related features of radiation ionization energies in semiconductors, J. appl. Phys. **39**, 2029-2038, 1968.

[11] U.S. Patent 4,121,933.

[12] H.-J. Hirsch, Properties of a radiographic system with lead oxide binder layers, J. photogr. Sci. **25**, 60-63, 1977.

[13] D. M. Korn, Influence of random fluctuations in charge transport upon image quality in electroradiography, J. appl. photogr. Engng. **4**, 48-51, 1978.

[14] S. Stotz, Particle charge and stability of liquid developers in electrophotography, in: W. F. Berg and K. Hauffe (ed.), Current problems in electrophotography, W. de Gruyter, Berlin 1972, 336-346.

the picture of fig. 14, and parallel to it, is constructed in the form of a sieve. This ensures that the greatest possible number of the lines of force leaving the charge pattern extend perpendicularly into the space occupied by the developer liquid, so that the field-strength there is at a maximum. The result is that the pigment is precipitated rapidly and over a wide area. In this way the image formed on the image carrier displays grey levels like those obtained in a conventional X-ray film photograph. Finally, the image has to be fixed; this can be done with a thin layer of lacquer.

Nature of the image

The use of the term 'picture processing' in the introduction has already given some indication that it is possible to change the nature of the image. A certain period of time is necessary for neutralizing the charge pattern and this time increases as the mobility and the concentration of the pigment particles and the ions in the developer liquid decrease. If development is stopped before the charge pattern has been completely neutralized, an image is obtained of the non-homogeneous fields present in the developing space as a result of the original charge distribution, because the charged particles mainly tend to follow closely spaced lines of force near transitions between adjacent image fragments. This favours the high spatial image frequencies, or in other words it is primarily contours that are enhanced. These are shown as a double line with a high and a low pigment concentration. This 'edge effect', which is not present in the latent image, is purely the result of the developing procedure. This case is called nonlinear development [15].

If we wait until the development is completed, in other words until the charge pattern has been completely neutralized, then the quantity of pigment that has been precipitated, and hence the total optical density of the image, is approximately proportional to the original local surface-charge density. This is called linear development. If moreover the surface charge is proportional to the X-ray exposure, then the entire radiographic system is linear: the nature of the image is the same as that obtained from the familiar silver film. In practice, however, the methods have to be based on a quasi-linear system and development is stopped before the charge pattern has been neutralized.

There are also other methods in which the nature of the image can be influenced in a very simple manner. It is possible to restore the contrast that has been lost owing to scattered radiation. Fig. 14 shows that a bias voltage can be applied between the carrier T and the developer electrode M during the development. Let us assume that a radiograph is being made of a simple object O as illustrated in *fig. 15*. Some of the radiation detected behind detail B originates from the scattered radiation from region A — a recurrent problem in radiography. In terms of intensity, the scattered radiation can assume values that are comparable with those of the primary radiation transmitted. The application of a bias voltage V during the development reduces the effective surface potential V_{eff} by the same amount for all picture elements (pixels). Fig. 15 illustrates the variation of the surface potential without (curve *1*) and with (curve *2*) a bias voltage. Curve *2* indicates some similarity to the situation without scattered radiation. The reduction in the surface potential (in a linear system) becomes evident through a reduction in the optical density of the image by a constant value for all pixels. This results in an improvement in the contrast (a real increase in the slope of the 'characteristic film curve').

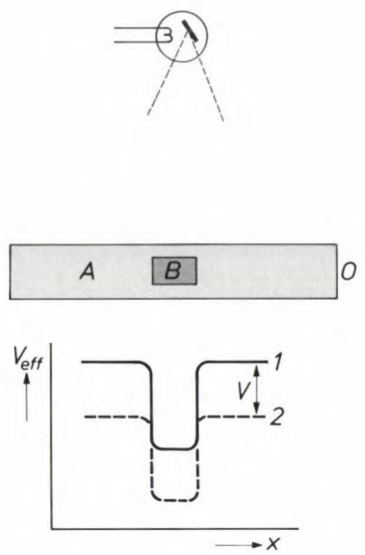


Fig. 15. Suppression of scattered radiation. When a radiograph is being made of object O, some of the radiation detected behind detail B originates from region A where this radiation has been scattered. The application of a bias voltage V to the developer electrode M in fig. 14 reduces the effective surface potential V_{eff} of the charge pattern by the same amount for all pixels (*dashed line*). This compensates for the contribution to V_{eff} caused by scattered radiation.

Limits to the sensitivity of development

The question now arises as to what is the development sensitivity that can be obtained. For a given charge density of the latent image it will be possible to obtain a higher optical density or blackening for a lower charge-to-mass ratio q/m of the separate particles of pigment. The charge q cannot however be made arbitrarily small — it cannot be smaller than the

electronic charge and the mass m may not be arbitrarily large because the precipitate would then have too coarse a grain.

A detailed analysis [16] shows that there is yet another stipulation in practice: development must not take longer than 30 to 60 s. This gives an optimum for the charge of the pigment particles; if q is too small, then too small a fraction of the image charges is neutralized within a given development time.

A further limit to the sensitivity occurs if the electrical force acting on the pigment becomes comparable with the forces produced by gravitation and the convection of the liquid. In practice this means that the development equipment also has to be optimized.

It is necessary to keep to a minimum the concentration of the ions in the developer liquid that have the same polarity as the pigment particles and compete with these particles in neutralizing the charge pattern. The required life and stability of the pigment suspension is a practical limitation to this [17].

Fig. 5*b* has shown us that, depending on the quality of the radiation, the surface-charge density of the charge pattern is normally $0.5\text{-}2.5 \times 10^{-9}$ C cm^{-2} mR^{-1}. To obtain an optical density of 2, a charge density of about 2×10^{-9} C/cm^2 is required for very sensitive development of grey levels. The required radiation dose for the foil (not to be confused with the much higher dose required for the object) for a high-pressure gas-ionography system of the type described is therefore less than 4 mR and greater than 0.8 mR. These values are comparable to film-foil systems with an average-to-high sensitivity.

Final comments

It has already been pointed out in the introduction that all electroradiographic systems have the common advantage that they do not use silver. Now that the principles of electroradiography in general and the ionography system in particular have been discussed in somewhat more detail, an attempt will be made to enumerate some of the other advantages and disadvantages of the electroradiographic method compared with methods employing silver film.

All electroradiographic methods now in existence have the disadvantage that they are insufficiently universal. It is not possible, for example, to make several exposures per second as it is with a film cassette changer. Tomography is difficult because of the following factors. The ionographic system is complicated by the thickness (1 cm) and the curved shape of the photoconductor. An acceptable lateral resolution can only be obtained when the electric lines of force between the electrodes in the chamber are directed towards the focus of the X-ray tube. Very complicated technical constructions would have to be made if ionigraphy were to be used for conventional tomography [18].

On the other hand, subtraction techniques are easier to apply than with film because the direction of movement of the charge carriers can be reversed. In pictures of blood vessels made with and without contrast medium only the differences between the two pictures are made visible, because the charge pattern that has been formed during the first exposure is partially or completely destroyed during the second. When the development has been completed, only the structures are visible in each image that have allowed radiation to pass through to varying degrees during the exposure.

Another special radiological method is mammography, but this has become the subject of debate in the last few years because of the risk of damage from the radiation. The requirements for picture quality are extremely high. These requirements can be met by (conventional) film mammography, because for film exposures the intensifying screen, which reduces the required radiation dose, is omitted and very soft (low-energy) radiation is used; this is largely absorbed in the tissue. Attempts — and these have often been critized — to use intensifying screens here mean that a compromise has to be made between reducing the radiation dose and reducing picture quality.

Xeroradiography is now being used for mammography, particularly in the United States. This can be done because selenium has a useful primary absorption at the low radiation energies to be used (see fig. 3, a maximum of about 80% effective absorption at about 26 keV). Nevertheless, the radiation dose is still very high because of the high conversion energy and the associated low sensitivity of the system. This cannot be compensated for by the particular electrophotographic method used for development.

Gas ionography is also being tried out for mammography. A phantom model has been used to determine the dose incident on the radiation-sensitive parenchyma for three systems: xeromammography, film-foil combination and mammography by ionography. The doses for the three systems have

[15] More details on liquid developing as a function of time are given in: H.-G. Junginger and R. Strunk, Latent image changes during electrophotographic development, J. appl. Phys. 47, 3021-3027, 1976.

[16] R. Kohler, D. Giglberger and F. Bestenreiner, Studies on electrophoretic developers for pictorial electrophotography, Photogr. Sci. Engng. 22, 218-227, 1978.

[17] Offenlegungsschrift 28 26 127 (German Patent Office publication).

[18] U.S. Patent 4,057,728.

been shown to be in the ratio of 4:2:1 [19]. One of the the ionography system is that the composition of the developer liquid has to be closely controlled, as discussed in the previous section. Also, the focus of the X-ray tube must be at a particular distance from the gas chamber and focused on it accurately. This system is not therefore compatible with existing X-ray mammography units.

For the reasons given above, it is impossible to say at this moment to what extent electroradiography will take over from film-foil radiography. It should also be remembered that a method that is analogous to silver-film recording will not solve the troublesome problems of storing large numbers of X-ray photographs. Future developments that come to mind here include electronic recording methods, which offer the advantages of automated information processing and the transmission of X-ray photographs over large distances ('Computer Radiography', CR).

[19] E. P. Muntz, M. Welkowsky, E. Kaegi, L. Morsell, E. Wilkinson and G. Jacobson, Optimization of electrostatic imaging systems for minimum patient dose or minimum exposure in mammography, Radiology **127**, 517-523, 1978.
L. Stanton, T. Villafana, J. L. Day, D. A. Lightfoot and R. E. Stanton, A study of mammographic exposure and detail visibility using three systems: Xerox 125, Min-R, and Xonics XERG, Radiology **132**, 455-462, 1979.

Summary. Economy in use of silver and new methods of development motivated studies to find out whether electroradiographic systems are suitable for use in medical X-ray diagnosis. Two electroradiographic systems — the high-pressure ionography system and the lead-oxide/gas-layer system — are discussed in detail, with a description of the formation of the latent charge pattern. At the Philips research laboratories in Aachen a special investigation of the sensitivity and resolution of the two systems has been made. In the section on the development procedure (making the latent charge pattern visible) consideration is also given to the possibility of 'picture processing' and the composition of the developer. Finally, the advantages and disadvantages of electroradiographic methods are compared with those of conventional radiography with silver films.

Scientific publications

These publications are contributed by staff of laboratories and plants which form part of or cooperate with enterprises of the Philips group of companies, particularly by staff of the following research laboratories:

Philips Research Laboratories, Eindhoven, The Netherlands *E*
Philips Research Laboratories, Redhill, Surrey, England *R*
Laboratoires d'Electronique et de Physique Appliquée, 3 avenue Descartes, 94450 Limeil-Brévannes, France *L*
Philips GmbH Forschungslaboratorium Aachen, Weißhausstraße, 51 Aachen, Germany *A*
Philips GmbH Forschungslaboratorium Hamburg, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany *H*
Philips Research Laboratory Brussels, 2 avenue Van Becelaere, 1170 Brussels (Boitsfort), Belgium *B*
Philips Laboratories, N.A.P.C., 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A. *N*

B. Aldefeld: Electromagnetic field diffusion in ferromagnetic materials.
Proc. IEE **125**, 278-282, 1978 (No. 4). *H*

E. Arnold & H. Schauer: Measurements of interface state density in MNOS structures.
Appl. Phys. Letters **32**, 333-335, 1978 (No. 5). *N*

H. Bacchi: Coding strategy for digital transmission of TV pictures at very reduced bit rates.
Digital image processing and analysis, ed. J. C. Simon & A. Rosenfeld, pp. 459-463; Noordhoff, Alphen aan den Rijn 1977. *L*

J. Bakker & G. A. Acket: Single-pass gain measurements on optically pumped $Al_xGa_{(1-x)}As$-$Al_yGa_{(1-y)}As$ double-heterojunction laser structures at room temperature.
IEEE J. **QE-13**, 567-573, 1977 (No. 8). *E*

N. M. Beekmans: Effect of oxygen chemisorption and photodesorption on the conductivity of ZnO powder layers.
J. Chem. Soc., Faraday Trans. I, **74**, 31-45, 1978 (No. 1). *R/E*

V. Belevitch, G. C. Groenendaal & R. R. Wilson (NKF Kabel, Delft): On the computation of primary parameters of circuits in twisted multiwire cables.
Proc. 26th Int. Wire and Cable Symp., Cherry Hill, N.J., 1977, pp. 338-342; 1978. *B, E*

A. Bensasson, A. Rabier & P. Harrop: Novel optimisation procedure applied to the minimisation of the ripple of a microwave amplifier.
Electronics Letters **13**, 685-686, 1977 (No. 22). *L*

M. Berth, M. Cathelin & G. Durand: Self-aligned planar technology for GaAs integrated circuits.
Tech. Dig. 1977 Int. Electron Devices Meeting, Washington D.C., pp. 201-204. *L*

S. Boronkay, B. Cassanhiol & J. Deforges (Faculté des Sciences, Poitiers): Couches minces de CdS photosensibles par pulvérisation cathodique RF réactive.
Le Vide **32**, 94-98, 1977 (No. 188). *L*

M. R. Boudry: An automatic system for broadband complex-admittance measurements on MOS structures.
J. Physics E **11**, 237-247, 1978 (No. 3). *R*

M. Boulou & D. Bois: Cathodoluminescence measurements of the minority-carrier lifetime in semiconductors.
J. appl. Phys. **48**, 4713-4721, 1977 (No. 11). *L*

P. W. J. M. Boumans: Erwiderung zur Stellungnahme von R. Plesch zu meiner Arbeit in dieser Zeitschrift, **279**, 1-16 (1976) *(Einige Überlegungen zur Situation der simultanen Multielementanalyse von Lösungen).*
Fresenius Z. anal. Chemie **288**, 64, 1977 (No. 1/2). *E*

P. W. J. M. Boumans & F. J. de Boer: An experimental study of a 1-kW, 50-MHz RF inductively coupled plasma with pneumatic nebulizer, and a discussion of experimental evidence for a non-thermal mechanism.
Spectrochim. Acta **32B**, 365-395, 1977 (No. 9/10). *E*

J. P. Boutot & M. Audier: Contribution des multiplicateurs à galette de microcanaux à la conception de nouveaux détecteurs.
C.R. Journées de Saclay sur les Détecteurs et ensembles de détection utilisés en physique nucléaire, 1977, pp. 17-38. *L*

J. P. Boutot, J. C. Delmotte, J. A. Miehé* & B. Sipp* (* CRN et Université Louis Pasteur, Strasbourg): Impulse response of curved microchannel plate photomultipliers.
Rev. sci. Instr. **48**, 1405-1407, 1977 (No. 11). *L*

D. J. Breed, W. T. Stacy, A. B. Voermans, H. Logmans & A. M. J. van der Heijden: New bubble materials with high peak velocity.
IEEE Trans. **MAG-13**, 1087-1091, 1977 (No. 5). *E*

J. C. Brice: The cracking of Czochralski-grown crystals.
J. Crystal Growth **42**, 427-430, 1977. *R*

J. W. Broer: On two active-reader stimulants: multiple titles and inverse writing — maximizing a figure of merit for your publication.
J. tech. Writ. Comm. **7**, 151-157, 1977 (No. 2). *E*

J. Bruinink, C. G. A. Kregting & J. J. Ponjeé: Modified viologens with improved electrochemical properties for display applications.
J. Electrochem. Soc. **124**, 1854-1858, 1977 (No. 12). *E*

J. Bruinink & P. van Zanten: The response of an electrochromic display with viologens on a potential step.
J. Electrochem. Soc. **124**, 1232-1233, 1977 (No. 8). *E*

R. Bruno, W. Hermann, H. Hörster, R. Kersten & F. Mahdjuri: Analysis and optimization of solar hot water systems.
Energy use management, Proc. int. Conf., Tucson 1977, Vol. 1, pp. 545-554. *A*

K. Bulthuis: Optical video-disc systems.
Laser 77 Opto-electronics, Proc. Conf. Munich 1977, pp. 637-642. *E*

E. Bunge, U. Höfker*, P. Jesorsky*, B. Kriener* & D. Wesseling* (* Heinrich-Hertz-Institut für Nachrichtentechnik, Berlin): Statistical techniques for automatic speaker recognition.
1977 IEEE Int. Conf. on Acoustics, speech & signal processing, Hartford, Conn., pp. 772-775. *H*

A. L. J. Burgmans, M. F. H. Schuurmans & B. Bölger: Transient behavior of optically excited vapor atoms near a solid interface as observed in evanescent wave emission.
Phys. Rev. A **16**, 2002-2007, 1977 (No. 5). *E*

C. M. van der Burgt, G. A. van Maanen & J. W. Vegt: Equipment for measuring piezoelectric polymer films in a wide frequency range: data on polyvinylidene fluoride PVDF.
Ultrasonics International 1977, Proc. Conf. Brighton, pp. 500-507. *E*

K. H. J. Buschow: Magnetic properties of borides.
Boron and refractory borides, ed. V. I. Matkovich, pp. 494-515; Springer, Berlin 1977. *E*

K. H. J. Buschow: Intermetallic compounds of rare-earth and 3d transition metals.
Rep. Prog. Phys. **40**, 1179-1256, 1977 (No. 10). *E*

K. H. J. Buschow, R. L. Cohen* & K. W. West* (* Bell Laboratories, Murray Hill, N.J.): What is the mechanism of hydrogen absorption in rare-earth intermetallics?
J. appl. Phys. **48**, 5289-5295, 1977 (No. 12).. *E*

K. H. J. Buschow & R. C. Sherwood (Bell Laboratories, Murray Hill, N.J.): Magnetic properties and hydrogen absorption in rare-earth intermetallics of the type RMn_2 and R_6Mn_{23}.
J. appl. Phys. **48**, 4643-4648, 1977 (No. 11). *E*

F. M. A. Carpay: The effect of pore drag on ceramic microstructures.
Ceramic microstructures '76, ed. R. M. Fulrath & J. A. Pask, pp. 261-275; Westview Press, Boulder, Colo., 1977. *E*

V. H. C. Evers & N. Hazewindus: Electronic media and home-based learning.
Aspects of Educational Technology **XI**, 209-216, 1977. *E*

J. B. Coughlin, J. B. Hughes, R. S. Watts & F. W. Siegert: A high speed ECL multiplexer in beam lead, hybrid technology.
Electrocomponent Sci. Technol. **4**, 185-191, 1977 (No. 3/4). *R, E*

J. E. Curran: The preparation of surface layers for electronic components by vacuum-deposition methods.
Advances in surface coating technology, Proc. int. Conf., London 1978, pp. 189-199. *R*

R. Davies & B. H. Newton: Microwave hybrid integrated circuit technology.
Wireless World **84**, Feb. 1978, pp. 46-50 + 67 (No. 1506). *R*

E. H. L. J. Dekker, K. L. L. van Mierloo & R. de Werdt: Combination of field and current access magnetic bubble circuits.
IEEE Trans. **MAG-13**, 1261-1263, 1977 (No. 5). *E*

M. Delfino, G. M. Loiacono & J. A. Nicolosi: Halide effect in L(+) glutamic acid halogen acid salts.
J. solid State Chem. **23**, 289-296, 1978 (No. 3/4). *N*

P. Delsarte: Hahn polynomials, discrete harmonics, and *t*-designs.
SIAM J. appl. Math. **34**, 157-166, 1978 (No. 1). *B*

P. Delsarte, J. M. Goethals & J. J. Seidel (Eindhoven University of Technology): Spherical codes and designs.
Geometriae Dedicata **6**, 363-388, 1977 (No. 3). *B*

P. A. Devijver: Decision-theoretic and related approaches to pattern classification.
Pattern recognition: theory and application, ed. K. S. Fu & A. B. Whinston, pp. 1-34; Noordhoff, Leiden 1977. *B*

A. M. van Diepen, H. J. Vledder & C. Langereis: The nature of the passivating oxide layer on iron powder.
Appl. Phys. **15**, 163-166, 1978 (No. 2). *E*

C. Dierieck: Some contributions to best approximation in normed linear spaces.
Bull. Soc. Math. Belg. **29**, 97-107, 1977 (No. 1). *B*

A. G. Dirks & H. J. Leamy: Columnar microstructure in vapor-deposited thin films.
Thin Solid Films **47**, 219-233, 1977 (No. 3). *E*

A. Dorsman, O. Elgersma, G. Meijer & L. Steuten: Chlorophyll metabolism in etiolated gherkin seedlings I. Photoinhibition of chlorophyll accumulation.
Photochem. Photobiol. **26**, 533-539, 1977 (No. 5). *E*

H. Durand & G. Pouvesle: Les Laboratoires d'Electronique et de Physique appliquée (LEP).
Rev. ANDES No. 23, 20-25, 1977. *L*

W. G. Essers, G. W. Tichelaar & G. A. M. Willems: Plasma-MIG-Schweißen von Aluminium.
Aluminium **53**, 663-666, 1977 (No. 11). *E*

V. H. C. Evers & N. Hazewindus: Electronic media and home-based learning.
Aspects of Educational Technology **XI**, 209-216, 1977. *E*

L. F. Feiner: Single ion and cooperative Jahn-Teller effect for a nearly degenerate E doublet.
Electron-phonon interactions and phase transitions, ed. T. Riste, pp. 345-350; Plenum Press, New York 1977. *E*

I. Flinn & J. D. Downes: The electromechanical properties of piezoelectric ceramics at high drive levels.
Ultrasonics International 1977, Proc. Conf. Brighton, pp. 145-153. *R*

M. J. C. van Gemert & T. G. Verbeek (Philips Lighting Division, Eindhoven): Instability of a low-pressure Na–noble-gas discharge.
Appl. Phys. Letters **31**, 500-502, 1977 (No. 8). *E*

J. Georges & F. Heymans: On parametric array expressions.
SIGPLAN Not. **13**, No. 1, pp. 43-47, Jan. 1978. *B*

A. Gerbier (collective pseudonym of an AFCET working-party including **M. Sintzoff**): Mes premières constructions de programmes.
Lecture Notes in Computer Science **55**, XII + 256 pp., 1977. *B*

H. M. Gibbs (Bell Laboratories, Murray Hill, N.J.), **Q. H. F. Vrehen & H. M. J. Hikspoors:** Single-pulse superfluorescence in cesium.
Phys. Rev. Letters **39**, 547-550, 1977 (No. 9). *E*

W. Gissler (Euratom, Ispra, Italy) **& R. Memming:** Photoelectrochemical processes at semiconducting WO_3 layers.
J. Electrochem. Soc. **124**, 1710-1714, 1977 (No. 11). *H*

J. J. Goedbloed: Comments on 'Avalanche buildup time of silicon reach-through photodiodes'.
J. appl. Phys. **48**, 4004-4005, 1977 (No. 9). *E*

R. G. Gossink: The preparation of ultrapure glasses for optical waveguides.
J. non-cryst. Solids **26**, 112-157, 1977 (No. 1-3). *E*

J. Hallais, D. Boccon-Gibod, J. P. Chané, J. M. Durand & L. Hollan: Device quality GaAs grown at low temperature by the halide process.
J. Electrochem. Soc. **124**, 1290-1294, 1977 (No. 8). *L*

P. Hansen, W. Tolksdorf & R. Krishnan (C.N.R.S., Bellevue, France): Anisotropy and magnetostriction of cobalt-substituted yttrium iron garnet.
Phys. Rev. B **16**, 3973-3986, 1977 (No. 9). *H*

W. M. A. Hax & W. S. M. Geurts van Kessel (State University of Utrecht): High-performance liquid chromatographic separation and photometric detection of phospholipids.
J. Chromatography **142**, 735-741, 1977. *E*

J. P. Hazan, J. J. Bernard & D. Küppers: Medium-numerical-aperture low-pulse-dispersion fibre.
Electronics Letters **13**, 540-542, 1977 (No. 18). *L, A*

D. Hennings & H. Schreinemacher: Temperature dependence of the segregation of calcium titanate from solid solutions of $(Ba,Ca)(Ti,Zr)O_3$ and its effect on the dielectric properties.
Mat. Res. Bull. **12**, 1221-1226, 1977 (No. 12). *A*

A. van Herk: Analytical expressions for side fringing response and crosstalk with finite head and track widths.
IEEE Trans. **MAG-13**, 1764-1766, 1977 (No. 6). *E*

I. S. Herschberg (Unilever, Rotterdam) **& J. C. A. Boekhorst:** Concurrent file access under unpredictability.
Inform. Process. Letters **6**, 203-208, 1977 (No. 6). *E*

L. Heyne: Electrochemistry of mixed ionic-electronic conductors.
Solid electrolytes, ed. S. Geller (Topics in applied physics **21**), pp. 169-221; Springer, Berlin 1977. *E*

H. Hieber, F. Betke & K. Pape: Ageing tests on gold layers and bonded contacts.
Electrocomponent Sci. Technol. **4**, 89-94, 1977 (No. 2). *H*

W. Hoekstra & V. H. C. Evers: Automated microfiche filing system for personal use.
IEEE Trans. **PC-20**, 228-233, 1977 (No. 4). *E*

H. Ihrig: The phase stability of $BaTiO_3$ as a function of doped 3d elements: an experimental study.
J. Physics C **11**, 819-827, 1978 (No. 4). *A*

G. Jacob, M. Boulou & M. Furtado: Effect of growth parameters on the properties of GaN:Zn epilayers.
J. Crystal Growth **42**, 136-143, 1977. *L*

L. Jacomme: Effects of an abrupt central dip in CVD fabricated fibers.
Wave Electronics **3**, 169-176, 1977/78 (No. 2). *L*

A. R. Jamieson (New Zealand Post Office, Wellington) **& T. E. Rozzi:** Rigorous analysis of crosspolarisation in flange-mounted rectangular-waveguide radiators.
Electronics Letters **13**, 742-744, 1977 (No. 24). *E*

W. H. de Jeu: Molecular structure and the occurrence of smectic A and smectic C phases.
J. Physique **38**, 1265-1273, 1977 (No. 10). *E*

W. H. de Jeu & W. A. P. Claassen: The elastic constants of nematic liquid crystalline terminally substituted azoxybenzenes.
J. chem. Phys. **67**, 3705-3712, 1977 (No. 8). *E*

W. J. M. J. Josquin & Y. Tamminga: Rutherford backscattering study of crystal orientation dependent annealing effects in high-dose antimony implanted silicon.
Appl. Phys. **15**, 73-78, 1978 (No. 1). *E*

R. A. J. Keijser: Polarization properties of internal mirror He-Ne lasers in a strong transverse magnetic field.
Optics Comm. **23**, 194-198, 1977 (No. 2). *E*

J. L. W. Kessels: A conceptual framework for a non-procedural programming language.
Comm. ACM **20**, 906-913, 1977 (No. 12). *E*

J. T. Klomp: Effekt van de mikrostruktuur van Al_2O_3-keramiek op het verbinden van het keramiek aan metaal.
Klei en Keramiek **27**, 170-182, 1977 (No. 10). *E*

K. Klose: Application of additional mirrors for rectilinear laser scanning of wide formats.
Appl. Optics **17**, 203-210, 1978 (No. 2). *H*

A. G. Knapp & J. R. Hughes: The effect of electron bombardment on evaporated films of sodium aluminium fluoride.
Proc. 7th Int. Vacuum Congress & 3rd Int. Conf. on Solid Surfaces, Vienna 1977, Vol. III, pp. 2161-2163. *R*

A. J. R. de Kock, W. M. van de Wijgert, J. H. T. Hengst, P. J. Roksnoer & J. M. P. L. Huijbregts: Microdefects and striations in dislocation-free LEC-GaP crystals.
J. Crystal Growth **41**, 13-28, 1977 (No. 1). *E*

G. Kowalski: The influence of fixed errors of a detector array on the reconstruction of objects from their projections.
IEEE Trans. **NS-24**, 2006-2016, 1977 (No. 5). *H*

G. Kowalski: Reduction of circular artefacts in fan-beam scanners.
IEEE Trans. **NS-24**, 2555-2560, 1977 (No. 6). *H*

B. Kramer, M. Parisot & A. Collet: Modular construction of low-noise multi-stage f.e.t. amplifiers.
Radio and electronic Engr. **48**, 23-28, 1978 (No. 1/2). *L*

P. Kramer & L. J. van Ruyven (Philips Semiconductor Development Laboratory, Nijmegen): Space charge influence on resistivity measurements.
Solid-State Electronics **20**, 1011-1019, 1977 (No. 12).

D. J. Kroon: Air pollution and the glass industry.
J. non-cryst. Solids **26**, 414-455, 1977 (No. 1-3). *E*

J.-P. Krumme, I. Bartels, B. Strocka, K. Witter, Ch. Schmelzer* & R. Spohr* (* Gesellschaft für Schwerionenforschung, Darmstadt): Pinning of 180° Bloch walls at etched nuclear tracks in LPE-grown iron garnet films.
J. appl. Phys. **48**, 5191-5196, 1977 (No. 12). *E, H*

H. Kurz: Optical information storage in photorefractive $LiNbO_3$.
Laser 77 Opto-electronics, Proc. Conf. Munich 1977, pp. 656-665. *H*

H. Kurz, V. Doormann & R. Kobs: Photorefractive recording of volume holograms in doped $LiNbO_3$.
Applications of holography and optical data processing, ed. E. Marom, A. A. Friesem & E. Wiener-Avnear, pp. 361-370; Pergamon Press, Oxford 1977. *H*

F. J. Kylstra (Eindhoven University of Technology) **& C. H. Smedema:** Real time programming.
Journal **A 18**, 185-193, 1977 (No. 4). *E*

M. Lacroix & A. Pirotte: ILL: an English structured query language for relational data bases.
Architecture and models in data base management systems, pp. 237-260; North-Holland, Amsterdam 1977. *B*

H. de Lang & N. H. Dekkers: Microscope adapted for the generation of various contrast types: theory and experiment.
Appl. Optics **16**, 2215-2222, 1977 (No. 8). *E*

P. K. Larsen, N. V. Smith* & H. H. Farrell* (* Bell Laboratories, Murray Hill, N.J.): Angle-resolved photo-emission study of chlorine chemisorbed on cleaved silicon (111).
Proc. 5th Int. Conf. on Vacuum ultraviolet radiation physics, Montpellier 1977, Vol. II, pp. 241-243. *E*

F. H. de Leeuw: Physikalische Grundlagen der Magnetblasen-Speicher.
NTG-Fachber. **58**, 199-209, 1977. *E*

F. H. de Leeuw: Wall velocity in garnet films at high drive fields.
IEEE Trans. **MAG-13**, 1172-1174, 1977 (No. 5). *E*

F. H. de Leeuw: New developments in the dynamics of magnetic domain walls and bubbles.
J. Magn. magn. Mat. **6**, 183-185, 1977. *E*

G. M. Loiacono, W. N. Osborne & M. Delfino: Crystal growth and properties of $Cs(D_xH_{1-x})_2AsO_4$.
J. Crystal Growth **41**, 45-55, 1977 (No. 1). *N*

F. K. Lotgering & P. H. G. M. Vromans: Chemical instability of metal-deficient hexagonal ferrites with W structure.
J. Amer. Ceramic Soc. **60**, 416-418, 1977 (No. 9/10). *E*

P. J. Mabey: Addressing in radio paging.
Electronics Letters **13**, 772, 1977 (No. 25). *R*

R. Madar, D. Michel, G. Jacob & M. Boulou: Growth anisotropy in the GaN/Al_2O_3 system.
J. Crystal Growth **40**, 239-252, 1977 (No. 2). *L*

G. M. Martin, J. Hallais & G. Poiblaud: Study of experimental conditions in thermally stimulated current measurements: application to the characterization of semi-insulating chromium-doped gallium arsenide.
J. Electrostatics **3**, 223-232, 1977 (No. 1/2/3). *L*

D. Mateika, P. Flisikowski, H. Kohler & R. Kilian: Automatisierte Kristallziehanlage für das Czochralski-Verfahren.
J. Crystal Growth **41**, 262-274, 1977 (No. 2). *H*

D. Mateika & Ch. Rusche: Coupled substitution of gallium by magnesium and zirconium in single crystals of gadolinium gallium garnet.
J. Crystal Growth **42**, 440-444, 1977. *H*

R. Memming: A rotating platinum-semiconductor ring-disc electrode for analysis of photoelectrochemical processes.
Ber. Bunsen-Ges. phys. Chemie **81**, 732-735, 1977 (No. 8). *H*

R. Memming: The role of energy levels in semiconductor-electrolyte solar cells.
J. Electrochem. Soc. **125**, 117-123, 1978 (No. 1). *H*

R. Metselaar: Electromagnetic properties of ferrites.
Interaction of radiation with condensed matter, Vol. II, pp. 159-221; IAEA, Vienna 1977. *E*

R. Metselaar (Eindhoven University of Technology) **& M. A. H. Huyberts:** Nonstoichiometry and electronic defects in yttrium iron garnet.
J. solid State Chem. **22**, 309-319, 1977 (No. 3). *E*

D. Meyer-Ebrecht & W. Spiesberger: Neue Verfahren in der medizinischen Röntgendiagnostik — eine Herausforderung für die industrielle Röntgenprüftechnik?
Materialprüfung **19**, 409-415, 1977 (No. 10). *H*

A. Mircea, A. Mitonneau, J. Hallais & M. Jaros (University of Newcastle-upon-Tyne): Study of the main electron trap in $Ga_{1-x}In_xAs$ alloys.
Phys. Rev. B **16**, 3665-3675, 1977 (No. 8). *L*

A. Mitonneau, G. M. Martin & A. Mircea: Hole traps in bulk and epitaxial GaAs crystals.
Electronics Letters **13**, 666-668, 1977 (No. 22). *L*

A. Mitonneau, G. M. Martin & A. Mircea: Spectroscopie capacitive des niveaux profonds dans les semiconducteurs.
Séminaire sur les méthodes d'analyse et de caractérisation des couches minces, Les Arcs - Bourg-Saint-Maurice 1978 (Suppl. Le Vide No. 189), pp. 33-39. *L*

B. J. Mulder: A method for the splash-free evaporation of aluminium and other metals.
Vacuum **28**, 11-12, 1978 (No. 1). *E*

J. H. Neave & B. A. Joyce: Temperature range for growth of autoepitaxial GaAs films by MBE.
J. Crystal Growth **43**, 204-208, 1978 (No. 2). *R*

J. A. Pals: Frequency dependence of microwave-enhanced superconductivity in Al strips with and without local injection of electrons.
Physics Letters **63A**, 141-143, 1977 (No. 2). *E*

J. A. Pals & W. van Haeringen: On the Josephson effect between superconductors in singlet and triplet spin-pairing states.
Physica **92B**, 360-365, 1977 (No. 3). *E*

J. Paredaens & D. Wybaux: The translation of relational data base languages to hierarchical languages.
Jahrestagung Gesellschaft für Informatik, Nürnberg 1977, pp. 45-46. *B*

L. J. van der Pauw: The radiation of electromagnetic power by microstrip configurations.
IEEE Trans. **MTT-25**, 719-725, 1977 (No. 9). *E*

C. Piaget, J. Vannimenus (E.N.S., Paris) **& P. Saget:** Electron energy distributions from GaP with negative electron affinity: I. Emission properties of thermalized electrons, II. Near-uv photoemission and secondary emission.
J. appl. Phys. **48**, 3901-3906, 3907-3913, 1977 (No. 9). *L*

R. J. van de Plassche & R. E. J. van der Grift: A five-digit analog-digital converter.
IEEE J. **SC-12**, 656-662, 1977 (No. 6). *E*

A. Posthuma de Boer & A. J. Pennings (State University of Groningen): Polyethylene networks cross-linked in solution: preparation, elastic behavior, and oriented crystallization. 2. Stress-strain behavior.
Macromolecules **10**, 981-989, 1977 (No. 5). *E*

G. Renelt: Runlength coding systems for black-and-white facsimile pictures: a comparison.
Electronics Letters **13**, 573-575, 1977 (No. 19). *H*

W. Rey: Admissibilité des estimateurs de type M en régression robuste.
Colloque IRIA, Sept. 1977, pp. 303-310. *B*

J. M. Robertson, M. Jansen, B. Hoekstra & P. F. Bongers: Growth of spinel ferrite films by liquid phase epitaxy.
J. Crystal Growth **41**, 29-35, 1977 (No. 1). *E*

U. Rothgordt & G. Renelt: Intermediate ternary code: a redundancy-reducing runlength code for digital facsimile transmission.
Electronics Letters **13**, 749-750, 1977 (No. 24). *H*

E. Roza & P. W. Millenaar: An experimental 560 Mbits/s repeater with integrated circuits.
IEEE Trans. **COM-25**, 995-1004, 1977 (No. 9). *E*

T. Rozzi, T. Itoh* & L. Grun* (* University of Illinois, Urbana): Two-dimensional analysis of the GaAs double hetero stripe-geometry laser.
Radio Sci. **12**, 543-549, 1977 (No. 4). *E*

L. J. van Ruyven, H. J. A. Bluyssen*, R. W. van der Heijden*, T. B. Tan* & H. I. Ralph (* University of Nijmegen): Hall-effect determination of the N-trap bound state in $GaAs_{1-x}P_x$.
Appl. Phys. Letters **31**, 685-687, 1977 (No. 10). *E, R*

I. Sander: Reversibel digitale Datenaufzeichnung in magnetooptischen Schichten.
Laser 77 Opto-electronics, Proc. Conf. Munich 1977, pp. 643-649. *H*

I. Sander & E. Krätzig: Optische Datenspeicher.
Elektro-Anzeiger **30**, No. 21, pp. 28-32, 1977. *H*

B. Schiek & J. Köhler: A method for broad-band matching of microstrip differential phase shifters.
IEEE Trans. **MTT-25**, 666-671, 1977 (No. 8). *H*

B. Schiek & J. Köhler: Improving the isolation of 3-dB couplers in microstrip-slotline technique.
IEEE Trans. **MTT-26**, 5-7, 1978 (No. 1). *H*

B. Schiek, W. Schilz & T. Paukner: A measuring system for the industrial application of microwave spectroscopy.
J. Microwave Power **12**, 347-359, 1977 (No. 4). *H*

P. Schnabel: Dispersion of thickness vibrations of piezoceramic disk resonators.
IEEE Trans. **SU-25**, 16-24, 1978 (No. 1). *A*

J. Schwandt: An approach to use evaluation nets for the performance evaluation of transaction-oriented business computer systems.
Computer performance, ed. K. M. Chandy & M. Reiser, pp. 429-441; North-Holland, Amsterdam 1977. *H*

J. M. Shannon: *I-V* impurity profiling with a Schottky barrier.
Appl. Phys. Letters **31**, 707-709, 1977 (No. 10). *R*

L. A. Æ. Sluyterman (I, II), **O. Elgersma** (I) **& J. Wijdenes** (II): Chromatofocusing: isoelectric focusing on ion-exchange columns: I. General principles, II. Experimental verification.
J. Chromatography **150**, 17-30, 31-44, 1978 (No. 1). *E*

E. T. J. M. Smeets, J. Dieleman, F. H. M. Sanders & D. de Nobel: Passivation of silicon p-n junctions by slightly conductive chalcogenide films.
J. Electrochem. Soc. **124**, 1458-1459, 1977 (No. 9). *E*

N. V. Smith*, P. K. Larsen & S. Chiang* (* Bell Laboratories, Murray Hill, N.J.): Anisotropy of core-level photoemission from InSe, GaSe, and cesiated W(001).
Phys. Rev. B **16**, 2699-2706, 1977 (No. 6). *E*

W. T. Stacy, M. A. H. Huyberts, R. Metselaar & A. B. Voermans: Decomposition of garnet epitaxial layers caused by annealing.
J. Appl. Phys. **48**, 4766-4769, 1977 (No. 11). *E*

W. T. Stacy, M. A. H. Huyberts, R. Metselaar & A. B. Voermans: Decomposition of garnet epitaxial layers caused by annealing.
J. appl. Phys. **48**, 4766-4769, 1977 (No. 11). *E*

A. L. N. Stevels & F. Pingault (Massiot-Philips, Ailly-Le-Haut-Clocher, France): On the use of x-ray screens in mammography.
J. Electrochem. Soc. **124**, 1400-1404, 1977 (No. 9). *E*

R. Stevens, P. D. White, R. F. Mitchell, P. Moore* & M. Redwood* (* Queen Mary College, London): Stop-band level of 2-port SAW resonator filters.
Proc. 1977 Ultrasonics Symp., Phoenix (Arizona), pp. 905-908. *R*

S. Strijbos: Powder-wall friction: the effects of orientation of wall grooves and wall lubricants.
Powder Technol. **18**, 209-214, 1977 (No. 2). *E*

S. Strijbos, P. J. Rankin, R. J. Klein Wassink, J. Bannink & G. J. Oudemans: Stresses occurring during one-sided die compaction of powders.
Powder Technol. **18**, 187-200, 1977 (No. 2). *E*

A. L. Stuijts: Ceramic microstructures.
Ceramic microstructures '76, ed. R. M. Fulrath & J. A. Pask, pp. 1-26; Westview Press, Boulder, Colo., 1977. *E*

A. L. Stuijts: Basic and practical aspects in sintering nitrogen ceramics.
Nitrogen ceramics, ed. F. L. Riley, pp. 331-350; Noordhoff, Alphen aan den Rijn 1977. *E*

E. Stupp, A. Pelissier, M. Kidder & A. Milch: GaP negative-electron-affinity cold cathodes: a demonstration and appraisal.
J. appl. Phys. **48**, 4741-4748, 1977 (No. 11). *N*

J. van Suchtelen: Coarsening of eutectic structures during and after unidirectional growth.
J. Crystal Growth **43**, 28-46, 1978 (No. 1). *E*

A. Thayse & J.-P. Deschamps: Logic properties of unate discrete and switching functions.
IEEE Trans. **C-26**, 1202-1212, 1977 (No. 12). *B*

W. Tolksdorf: Growth of magnetic garnet single crystals from high temperature solution.
J. Crystal Growth **42**, 275-283, 1977. *H*

A. A. v. d. Veeke: Omzetters voor geschakelde voedingen.
Polytechn. T. Elektr. **32**, 602-612, 1977 (No. 11). *E*

A. W. Veenis & A. Bril: On vacuum-ultraviolet excitation spectra of luminescent substances.
Proc. 5th Int. Conf. on Vacuum ultraviolet radiation physics, Montpellier 1977, Vol. II, pp. 104-106. *E*

H. Verweij, H. van den Boom & R. E. Breemer: Raman scattering of carbonate ions dissolved in potassium silicate glasses.
J. Amer. Ceramic Soc. **60**, 529-534, 1977 (No. 11/12). *E*

J. Volger: Cryogenics: a critical review.
Interdiscipl. Sci. Rev. **2**, 312-326, 1977 (No. 4). *E*

C. Werkhoven, J. H. T. Hengst & W. J. Bartels: Effect of dislocation loops in macroscopically dislocation-free GaP substrates on the perfection of homo-epitactic deposits.
J. Crystal Growth **42**, 632-638, 1977. *E*

H. W. Werner: Comparison of thin film analytical methods.
Proc. 7th Int. Vacuum Congress & 3rd Int. Conf. on Solid Surfaces, Vienna 1977, Vol. III, pp. 2135-2144. *E*

C. H. Weysenfeld & M. M. M. P. Matthey: Scaling of positive-column Hg-Ar discharges towards display cells.
IEEE Trans. **ED-24**, 1239-1244, 1977 (No. 10). *E*

C. W. White (Oak Ridge National Laboratory, Oak Ridge, Tenn.), **E. W. Thomas** (Georgia Institute of Technology, Atlanta, Ga.), **W. F. van der Weg & N. H. Tolk** (Bell Laboratories, Murray Hill, N.J.): Optical emission from low-energy ion-surface collisions.
Inelastic ion-surface collisions, ed. N. H. Tolk, J. C. Tully, W. Heiland & C. W. White, pp. 201-252; Academic Press, New York 1977. *E*

J. C. Williams: A 36 GHz printed planar array.
Electronics Letters **14**, 136-137, 1978 (No. 5). *R*

J. P. Woerdman: A direct measurement of Na_2 Franck-Condon factors.
Chem. Phys. Letters **53**, 219-222, 1978 (No. 2). *E*

Published in Process and device modeling for integrated circuit design, ed. F. van de Wiele, W. L. Engl & P. G. Jespers, Noordhoff, Leiden 1977:

H. C. de Graaff: Review of models for bipolar transistors (pp. 283-306). *E*
H. C. de Graaff: High current density effects in the collector of bipolar transistors (pp. 419-442). *E*
H. C. de Graaff: Emitter effects in bipolar transistors (pp. 443-460). *E*
F. M. Klaassen: Survey of I^2L modelling (pp. 519-537). *E*
F. M. Klaassen: Review of physical models for MOS transistors (pp. 541-571). *E*
F. M. Klaassen: Characterization and measurements of MOST devices (pp. 573-588). *E*
F. M. Klaassen: A MOST model for CAD with automated parameter determination (pp. 739-750). *E*

PHILIPS TECHNICAL REVIEW

VOLUME 39, 1980, No. 2

A semiconductor laser for information read-out

J. C. J. Finck, H. J. M. van der Laak and J. T. Schrama

Semiconductor lasers only became efficient lasers when they were made in the form of 'double heterojunctions': an active layer in which the laser radiation is produced, between two layers of another composition. In the first lasers of this type the active layer consisted of GaAs. The laser radiation from this material has a wavelength of about 890 nm. Since then the development of semiconductor lasers has moved simultaneously in two opposing directions. On the one hand there is a clear need for laser radiation at longer wavelengths for optical communication by glass fibres, since Rayleigh scattering in the fibres is lower at these longer wavelengths. On the other hand there is a need for semiconductor lasers that produce radiation at shorter wavelengths for applications in which the radiated power is concentrated in the smallest possible area. This article describes a laser of this shorter-wavelength type that Philips is putting on the market. The wavelength of the emitted radiation is 780 nm.

Introduction

The information on a 'VLP' disc is recorded in a spiral track 25 km long, with a pitch of 1.66 μm and consisting of pits 0.6 μm wide and 0.12 μm deep. This form of information storage also has its attractions for applications such as audio discs and data filing, mainly because it can provide a very high information density — of the order of one bit per μm^2. Philips are active in the development of such systems ('Compact Disc'; 'DOR' = Digital Optical Recording).

The information stored in this way is read out by scanning the track of pits with an optical system; a photodiode converts the intensity variations of the reflected light into an electrical signal (*fig. 1*). The requirements imposed on the light source (*L*) for this system can only be met by lasers: the radiation must have a radiation density of about 2500 W per cm^2 per steradian for a detected signal with an acceptable signal-to-noise ratio to be generated by the variations in reflection from an area of about 1 μm^2 [1]. The light source in the 'VLP' system is at present a helium-neon laser [2].

J. C. J. Finck, Ir H. J. M. van der Laak and Ir J. T. Schrama are with the Philips Elcoma Division, Eindhoven.

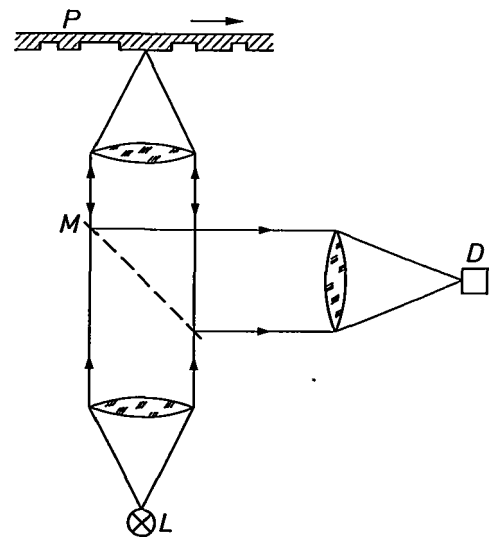


Fig. 1. Optical system for read-out from the track of pits on a 'VLP' disc. *L* laser. *M* half-silvered mirror. *P* 'VLP' disc. *D* photodetector.

[1] J. P. J. Heemskerk, Appl. Optics **17**, 2007, 1978.

[2] A description of the Philips 'VLP' system is given in the series of articles on this subject in Philips tech. Rev. **33**, 177-193, 1973, and Appl. Optics **17**, 1993-2042, 1978.

In this article we shall discuss an alternative to the HeNe laser in this application: the 'CQL10', a semiconductor laser based on the AlGaAs system. Lasers of this type have already proved to be highly suitable as light sources for optical communication by glass fibre [3]. A further development of this kind has now led to the CQL10.

The CQL10 differs greatly in many respects from the HeNe laser. For example, it does not produce a parallel beam, but a highly divergent beam, which requires a different optical system. The coherence length is relatively small; this is an advantage for the intended application; it reduces the chance of undesirable wave interference. Other advantages are that it is small, mechanically strong and only requires a low supply voltage. The HeNe laser is 20 cm long and requires a supply voltage of 1500 V, whereas the CQL10 measures about 1 cm and requires a voltage of 2 to 3 V. In quantity production the CQL10 will also be cheaper.

When we were designing the CQL10, the two main requirements were that it should be capable of faithful read-out from the existing 'VLP' discs and that it should work well even at a temperature of 60 °C. The last requirement relates to using the 'Compact Disc' in an ambient temperature of 45 °C (as in the tropics). Because of the compact construction of the unit the cooling capacity must necessarily be relatively small, so that the temperature of the encapsulation of the laser could rise as high as 60 °C.

The AlGaAs laser

The construction and principle of operation of a 'semiconductor laser with double heterojunction' will be outlined briefly with the aid of *fig. 2* [4]. The laser consists of three layers with different types of conductivity and different energy gaps E_g. In layers *1* and *3* E_g is greater than in the 'active' layer *2*.

A photon of the correct frequency ν ($h\nu \approx E_{g2}$) passing through layer *2* may generate a hole-electron pair or 'stimulate' an electron that may be present in the conduction band to recombine with a hole in the valence band. In a generation process the photon disappears ('absorption'), while in a recombination process a photon of the same phase is added ('stimulated emission').

If the semiconductor is in thermal equilibrium ($I = 0$ in *fig. 2a*), a current of such photons is strongly attenuated because there are few electrons in the conduction band and few holes in the valence band: virtually the only process that occurs is electron-hole generation (with absorption); recombinations (with emission) hardly occur at all. If there is an electric current (I) in the diode in the forward direction, however, electrons flow from layer *1* and holes from layer *3* into layer *2*. If this current is high enough the energy-level diagram takes the form shown in fig. 2*b*. As a result of the presence of the barriers ΔE_c and ΔE_v the electrons and holes collect in layer *2* and a *population inversion* arises, in which there are 'effectively' more occupied than unoccupied states in the conduction band, which means that on average the stimulated recombinations outnumber the excitations for every incident photon. A light wave that passes through this medium is amplified. This 'amplification by stimulated emission' is proportional to the 'excess' of electrons in the conduction band.

The grey region in fig. 2*a* is therefore an amplifier for light in the z-direction. By means of partially

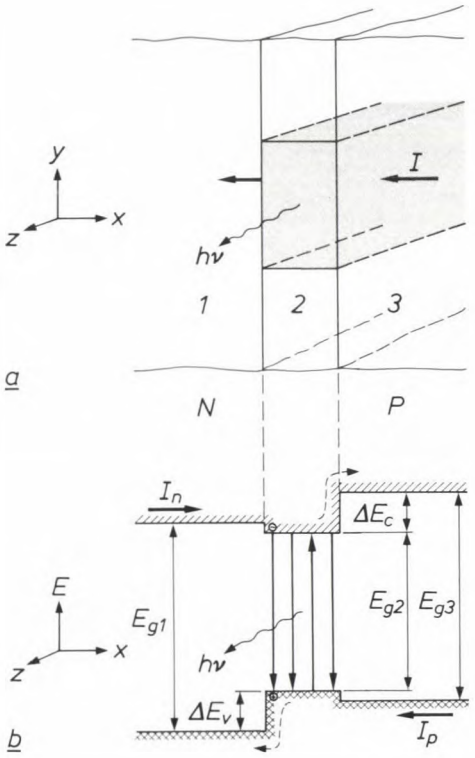


Fig. 2. Structure and principle of operation of the semiconductor laser with double heterojunction. *a*) The laser consists of three layers *1*, *2* and *3*; layer *1* has *N*-type conductivity, layer 3 has *P*-type conductivity and both have a larger energy gap than layer *2*, which is undoped (almost intrinsic). When a sufficiently large current (I) flows in the forward direction the grey part of layer *2* becomes 'active', i.e. becomes an amplifier for wavelengths of frequency ν ($h\nu \approx E_{g2}$). The semiconductor/air interfaces perpendicular to the z-axis at the front and back of the active region act as half-silvered mirrors, providing feedback in the amplifier. If the gain can compensate for the losses the resonator thus obtained will start to oscillate. 'Laser radiation' will then be emitted from the front and rear mirrors. *b*) A simplified energy-level diagram for a large value of the current I. Because of the presence of the barriers ΔE_c and ΔE_v the electrons and holes flowing towards layer *2* collect in this layer and readily cause a 'population inversion'. The layer is then active because the stimulated recombination processes with emission of coherent photons outnumber the generation processes with absorption. I_n is the electron current, I_p the hole current. E_{g1}, E_{g2}, E_{g3} are the energy gaps of the layers *1*, *2* and *3*.

transmitting mirrors at the front and back, feedback is provided and a resonator is produced. When the current I is gradually increased, the amplification on one pass through the resonator becomes greater at some instant than the internal losses and the losses at a mirror, and the resonator starts to oscillate: laser action occurs. The current at which this starts is known as the threshold current (I_{th}). For larger values of I the laser emits an intense coherent light beam from the two end mirrors. *Fig. 3* gives the light/current characteristic (optical power plotted against current) of the laser. For currents below the threshold the holes and electrons recombine *spontaneously,* with the emission of incoherent photons; the diode then behaves like an LED.

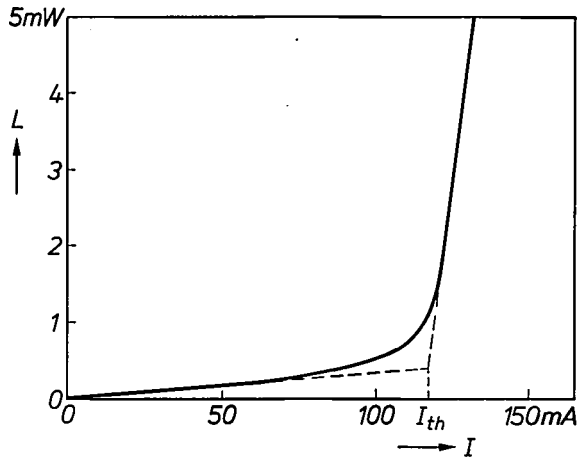


Fig. 3. Light/current characteristic of a laser, i.e. the power L of the emitted radiation as a function of the current I. Below the threshold I_{th} the diode acts as an LED (spontaneous emission) and above it as a laser (stimulated emission).

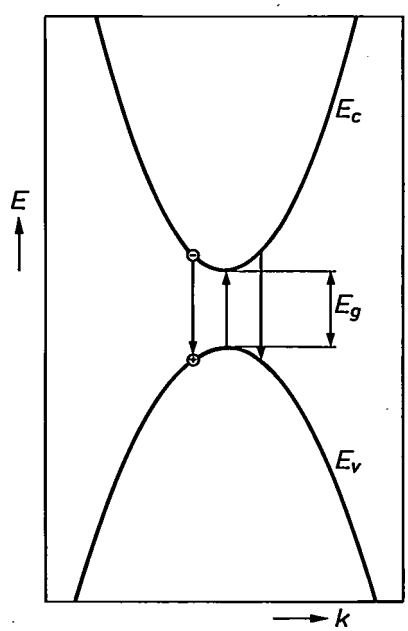


Fig. 4. Energy E as a function of the momentum (the wave number) k of the electrons in a 'direct' semiconductor. The minimum of the conduction band E_c and the maximum of the valence band E_v have the same value of k. Consequently an electron at the bottom of the conduction band can recombine 'directly' — i.e. without transfer of momentum to the lattice — with a hole at the top of the valence band. Because of this, the probability of a recombination or generation process for each incident photon is much greater than in 'indirect' semiconductors such as Ge and Si in which the maxima and minima of the bands are not directly opposite one another.

The secret of the 'semiconductor laser with double heterojunction' lies in the difference in energy gap between layer *2* and layers *1* and *3*. Not only does this create the barriers ΔE_c and ΔE_v (fig. 2*b*) that 'trap' the holes and electrons thus making it easy for a population inversion to occur, but it also creates a jump Δn in the refractive index that traps the photons in the x-direction. This is because an increase in the energy gap implies a reduction in the refractive index; layers *1* and *3* therefore have a lower refractive index than layer *2*, so that light waves travelling close to the z-direction are totally reflected at the interfaces.

The AlGaAs system — or to be more accurate the $Al_xGa_{1-x}As$ system — is particularly suitable for making a laser of this type. In the first place, GaAs is a 'direct' semiconductor (see *fig. 4*), which means that there is a high probability of absorption and stimulated emission. In the second place the Al ion in the zinc-blende lattice of $Al_xGa_{1-x}As$ is of about the same size as the Ga ion, so that layers of different composition can be epitaxially applied on top of one another without causing stresses. Because the energy gap and the refractive index have a marked dependence on x in accordance with the empirical relations [5]

$$E_g = 1.424 + 1.247x \text{ eV}, \quad (1)$$

$$n \approx 3.59 - 0.71x, \quad (2)$$

there is considerable freedom in the choice of these quantities. Relation (1) is only true for values of x between 0 and about 0.46. In this entire region the material is a direct semiconductor, outside this region it is not.

The front and back mirrors are in fact cleavage planes of the crystal and are consequently ideally plane and parallel. No special measures are taken to entrap the photons in the y-direction. However, no current is allowed to flow through layer *2* outside the active region (grey in fig. 2*a*), so that population inversion only occurs inside this region. Light waves in the z-direction derive the greatest effective gain from the inversion ('gain guiding').

[3] See for example G. A. Acket, J. J. Daniele, W. Nijman, R. P. Tijburg and P. J. de Waard, Philips tech. Rev. 36, 190, 1976.
[4] A detailed discussion is given in the article of note [3]; heterojunctions are also discussed in L. J. van Ruyven, A. Rev. Mat. Sci. 2, 501, 1972.
[5] H. C. Casey Jr. and M. B. Panish, Heterostructure lasers, Part A, pp. 45 and 192; Academic Press, London 1978.

The laser crystal of the CQL10

Fig. 5 gives a cross-section of the laser crystal of the CQL10. It consists of a substrate *S* with four layers *1*, *2*, *3*, *4*, which are applied to it by means of liquid-phase epitaxy (LPE) [6], and two metallic contact layers. The substrate and layer *4* are of pure GaAs. In the other layers some of the Ga has been replaced by Al: 16 at % in layer *2*, and 46 at % in layers *1* and *3*. We shall return later to the choice of these compositions, particularly for layers *1*, *2* and *3*. In the epitaxy, wafers of about 2 cm^2 are formed. These are cleaved into wafers of about 1 cm^2 which are then processed as described briefly below before being cleaved again into the individual laser crystals.

The thickness chosen for the active layer, 0.2 μm, is an optimum. As the thickness decreases, the threshold current generally decreases, because a population inversion has to be maintained in an ever-decreasing volume. However, if the layer becomes thinner than 0.2 μm, I_{th} increases again because the radiation is no longer effectively confined between layers *1* and *3*.

In the lateral direction (*y*) the current flow is limited to a stripe 5 μm wide, since the layers *3* and *4* outside this stripe have been made into insulators by proton implantation of the wafer (grey shading in fig. 5). During the implantation a metal mask protects the 5 μm stripes from the proton bombardment. The implantation front comes to within about 1 μm of the active layer. This provides adequate definition of the current flow, while disturbances of the crystal lattice in advance of the front still do not reach the active layer.

Pure GaAs has been chosen for the substrate (*S*) and the top layer (*4*), because Al at the surface oxidizes very readily and this makes it extremely difficult to apply good metal contacts. To help form good contacts the substrate is heavily *N*-doped with Si and the top layer is partly made into a degenerate *P*-semiconductor by a strong Zn diffusion. Before the substrate is metallized, it is ground and etched to a thickness of about 100 μm, to remove the Zn diffused into the substrate and make the cleavage easier.

A wafer is cleaved into about 30 'bars', each of which is broken into about 30 laser crystals, so that a wafer yields about 1000 lasers. The width of the bars, i.e. the distance between the mirror planes of the active region, is 250 μm.

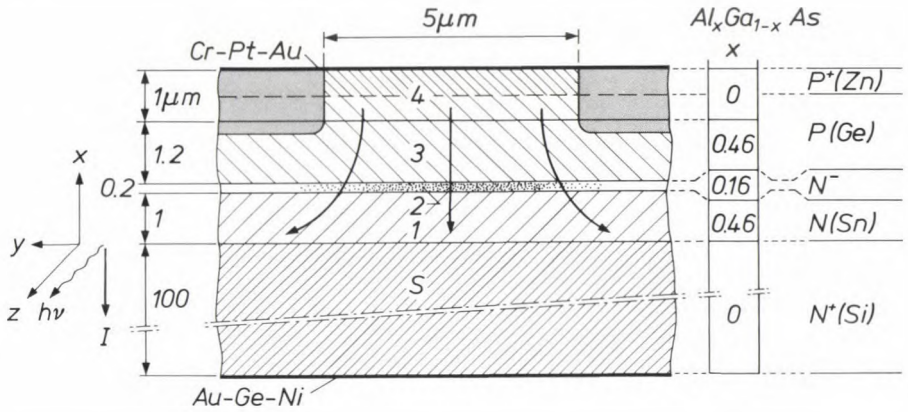


Fig. 5. Cross-section through the laser crystal of the CQL10. The thickness of the layers is given on the left and the Al content *x* (the composition is $Al_xGa_{1-x}As$) and the conduction type and dopant are given on the right. The undoped layer *2* is weakly *N*-type because of residual impurities. The upper part of layer *4* is doped to make it very strongly *P*-type (degenerated) by giving it a strong Zn diffusion. The upper layers are made into insulators by proton implantation (dark grey area), except for a narrow stripe (5 μm) that determines the passage of current and hence the width of the active region in layer *2*. The difference in composition between layer *2* and layers *1* and *3* provides the desired difference in energy gap. The heavy doping of the substrate *S* and the upper layer permits good electrical contact between the crystal and the metallic layers (Cr-Pt-Au and Au-Ge-Ni).

The composition of the active layer and the 'cladding' layers

As we shall now briefly explain, the selection of the values of 16% for the Al content x_2 of layer *2* and 46% for the Al content x_1 of layers *1* and *3* is directly linked to the requirement that this laser must be able to read the internationally standardized 'VLP' discs reliably even when the temperature of the laser encapsulation is 60 °C. In normal operation this means that the temperature of the active region is about 70 °C.

The wavelength of the HeNe laser now used for scanning 'VLP' discs is 630 nm. If a laser with a

longer wavelength is to be used, the permissible variations in disc thickness become smaller. For a GaAs laser ($x_2 = 0$, $\lambda = 890$ nm) the tolerances would become unacceptably small. An analysis has shown that the wavelength can be increased to about 780 nm for standard 'VLP' discs [7]. This wavelength corresponds to an energy gap of 1.62 eV, and hence to $x_2 = 0.16$ for an AlGaAs laser (see eq. 1). Since λ increases as x_2 decreases, this is a minimum value for x_2.

If the laser in a given 'VLP' unit is replaced by a laser of longer wavelength, the resolution decreases; the diameter (d) of the light spot increases. This can be compensated for by increasing the numerical aperture A ($d = 1.2\ \lambda/A$). However, this reduces the tolerance for the disc thickness. On the face of it this would be expected: the tolerance for a 'blunt' beam (large value of A) is smaller than for a sharp beam (small value of A). The relationship, however, is more complicated than this because of refraction effects at the edges of the pits. For standard 'VLP' discs the analysis gives a maximum of 780 nm for the wavelength, as mentioned above.

The use of a longer wavelength also means that crosstalk between adjacent tracks occurs more readily because of coma of the light spot. This problem arises because of 'sag' at the edge of the disc, which has the effect that the axis of the laser optical system is not always perpendicular to the disc. This problem can in principle be avoided by using an arrangement in which the optical axis follows the curvature of the disc.

To obtain high barriers and a large jump in the refractive index, the difference Δx between x_1 and x_2, and hence x_1, should be made as large as possible. The maximum temperature at which the probability of holes and electrons escaping across the barriers is still sufficiently small would then be as high as possible. However, if x_1 exceeds the value 0.46, there is a fundamental change in the band structure [8], which has an important consequence for the laser design: the depth of the donor levels (Sn) in layer *1* increases suddenly. This considerably reduces the concentration of electrons at the junction and injection suddenly becomes much more difficult. We note that raising x_1 above 0.46 would in any case not be very effective in raising the height of the barriers because equation (1) is no longer valid and the energy gap only increases slowly for $x_1 > 0.46$. For our laser design 0.46 is therefore a practical maximum for x_1.

With the jump in x thus limited to a value of 0.30, the probability of charge carriers escaping thermally from the active layer is still sufficiently small even at a somewhat higher temperature. An increased escape probability means a higher threshold current I_{th}. With $\Delta x = 0.30$, the threshold current I_{th} does not increase by more than 30% if the temperature rises 30 degrees. For a typical laser with a threshold current of 120 mA at 30 °C, the threshold current therefore remains below 160 mA at 60 °C, a very acceptable value. The value $\Delta x = 0.30$ also implies a jump of 0.2 in the refractive index, which is sufficient to trap the light in the x-direction.

Assembly and encapsulation of the crystal

To keep the manufacturing costs of the CQL10 as low as possible, we tried to make good use of well-proven standard components for the design of the encapsulation. *Fig. 6* is a diagram of the encapsulation. It consists of an assembly block *B* specially made for the laser, and two standard components: the baseplate *P* and the cap with window *C*. A *PIN* photodiode (*D*) of silicon attached to the baseplate is used for stabilizing the optical power from the laser crystal by feedback to the power supply. *Fig. 7* shows the components and the assembled laser.

The assembly block (fig. 6*b*) enables the laser crystal (*L* in fig. 6*a*) to be accurately located in the equipment. It is mainly cylindrical and is made with very close tolerances, particularly for the outside diameter (8.991-9.000 mm) and the relative position of the mounting surface (*A*) for the crystal. Two slots in its base permit the block to be accurately oriented in the equipment. It also functions as a heat sink for the heat generated in the crystal. (The electrical energy supplied is converted almost entirely into heat: the optical efficiency is 1 to 5%). The shape and

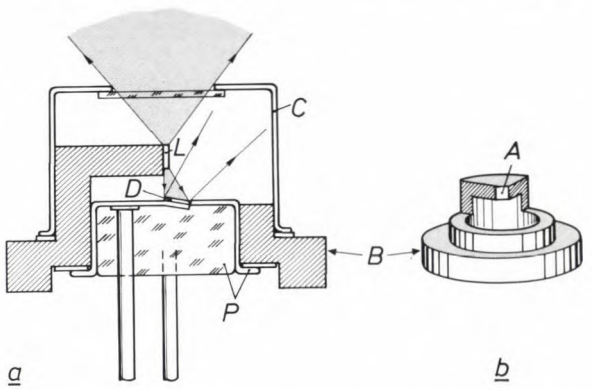


Fig. 6. *a*) Diagram of the encapsulation of the laser crystal. *P* baseplate, *B* assembly block, *C* cap with window. In addition to the laser crystal *L*, the assembled laser includes a photodiode *D*, which can be used to measure the radiation level. For clarity the three connecting pins (for chassis, laser crystal and photodiode) are not all shown. *b*) Diagram of the assembly block; the laser crystal is soldered to the face *A*.

[6] T. G. J. van Oirschot, W. J. Leswin, P. J. A. Thijs and W. Nijman, J. Crystal Growth **45**, 262, 1978.

[7] This analysis was carried out by P. F. Greve and Dr J. P. J. Heemskerk of the Philips Audio Division, and Ing. W. G. Ophey of Philips Research Laboratories, Eindhoven.

[8] See for example R. Dingle, R. A. Logan and J. R. Arthur Jr., in: Gallium arsenide and related compounds (Edinburgh) 1976 (Inst. Phys. Conf. Ser. No. 33a), p. 210.

material (oxygen-free copper) of the block are chosen so as to minimize its thermal resistance, which is about 5 kW.

When the crystal is being attached it must not be exposed to mechanical stresses resulting from differences in the expansion coefficients of crystal and block. It is therefore soldered with indium, which is very ductile and has a melting point of 156 °C.

The crystal is soldered with a special soldering jig to a tolerance of ± 50 μm in the transverse direction, ± 1° in the angle to the vertical and 4 μm in the vertical direction. The vertical tolerance is so small for the following reasons: firstly, the exit mirror must not lie below the edge of the block because the beam from affecting the laser or emerging within the aperture angle of interest in the application.

The window in the cap (*C* in fig. 6) must not affect the emergent wavefront. It has been found that a standard microscope cover glass meets the requirements for application in 'VLP' equipment, i.e. that its flatness must be such that it will give no more than two Newton's rings over the diameter. The window is hermetically sealed into the cap. Interference measurements are made to test the windows for flatness both before and after the cap is welded to the block. To protect the crystal, the encapsulation is filled with dry nitrogen (max. 100 ppm H_2O). A layer of gold (2 μm) protects the encapsulation from oxidation.

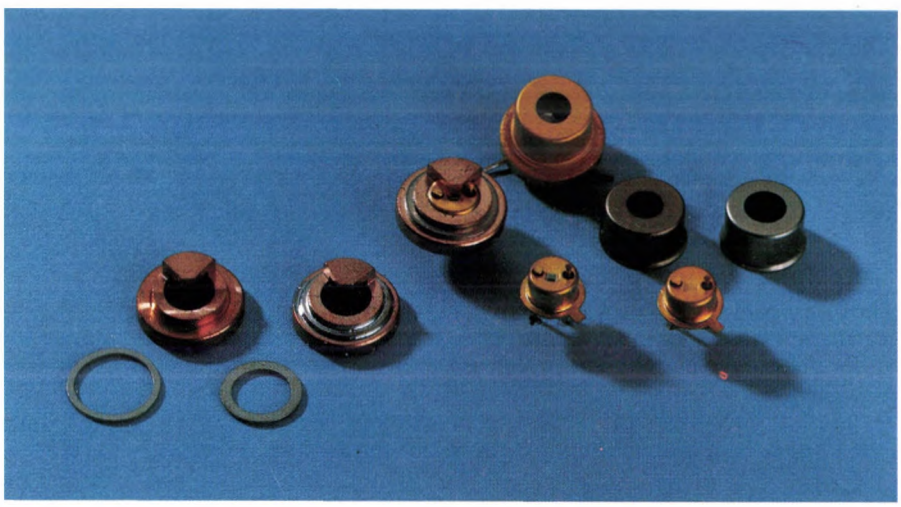


Fig. 7. The encapsulation components and the assembled laser. The various components and the assembled laser can be identified from fig. 6. The second assembly block from the left has the welding rings for baseplate and cap.

then reflected by it would interfere with the main beam, giving inadmissible variations in the intensity some distance away from the laser as a function of the angle to the optical axis. Secondly, the crystal should also not project more than 4 μm beyond the edge because the cooling of the mirror would then be inadequate. A great deal of heat is generated at the mirror itself because the cleavage plane is a crystallographic 'injury' that induces substantial non-radiative heat-generating recombination processes.

The baseplate has three connecting pins sealed in glass, for connection to the chassis, the laser crystal and the photodiode. It is a standard Philips product, the TO-18, which can have up to seven pins. The photodiode is soldered in an indentation in the baseplate at an angle of 10° to prevent the reflected light

Characteristics and use of the CQL10

Fig. 8 shows an example of the spectrum of the laser for a current above the threshold value. As noted earlier, the laser can be considered as an amplifier of light waves in the z-direction (see figs 2 and 5), which has become a resonator because of the mirrors. The envelope of the peaks in fig. 8 gives the gain profile of the amplifier (width about 3 nm). The peaks represent longitudinal modes of the resonator cavity; they occur where the resonance condition $l = m\lambda/2n$ is satisfied (l is the length of the resonator, m an integer, λ the wavelength of the light, n the refractive index of the resonator material). With $l = 250$ μm, $\lambda = 780$ nm, $n = 3.59$ the number of half-wavelengths m in the resonator will be about 2300 and the distance $\Delta\lambda$ between two adjacent peaks ($\Delta m = 1$) will be

about 0.3 nm, as in fig. 8. Closer inspection reveals that within the gain profile $\Delta\lambda$ increases significantly with λ, owing to the dependence of n on λ (i.e. dispersion). The width of the envelope and the number of peaks decrease as the current increases. At a high current there is clearly one dominant mode. A spectrum with many modes implies a short coherence length.

The exit beam from this laser, unlike that from the HeNe laser, is strongly divergent. This is because the exit aperture is of the order of magnitude of the wavelength, so that there is considerable refraction. *Fig. 9* gives the measured intensity as a function of the angle to the optical axis both in the plane of the active layer and perpendicular to it. The aperture of the laser is least, and the divergence therefore greatest, in the plane perpendicular to the active layer. The peak at $-20°$ in fig. 9 is from light that is reflected by the photodiode, which is at an angle of 10° (page 42); it falls completely outside the angle of aperture of interest for read-out applications.

In addition, the laser light is 'astigmatic': the emergent wavefront is more strongly curved in the x-direction than in the y-direction (*fig. 10a*). The centre of curvature for the x-direction lies in the exit mirror, and the centre of curvature for the y-direction is about 25 μm behind it in the laser. Astigmatic laser radiation is a characteristic of lasers with 'gain guiding' (page 39) in the lateral direction (y in fig. 10). The gradual transition from 'amplification' to 'absorption' in the y-direction at the edge of the active region is linked with a decrease in this same direction in the velocity of light (i.e. with an increase in the refractive index). The light therefore lags somewhat at the edge so that the wavefront is already curved inside the laser (fig. 10*b*). After refraction at the exit plane,

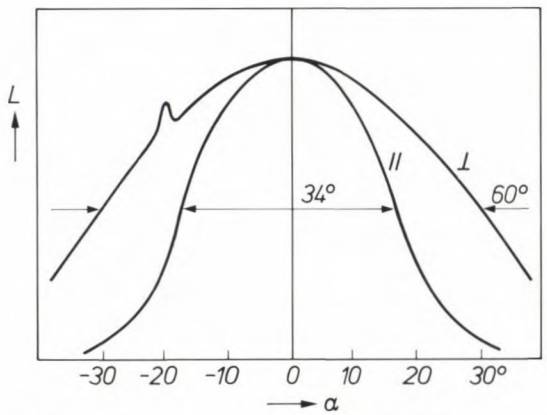


Fig. 9. The power L of the emitted radiation as a function of the angle α to the optical axis of the laser, for radiation in the plane of the active layer (||) and for radiation in the plane perpendicular to it (⊥). The peak at $-20°$ is caused by reflection at the photodiode. The 'half-power angles' are 34° and 60° respectively.

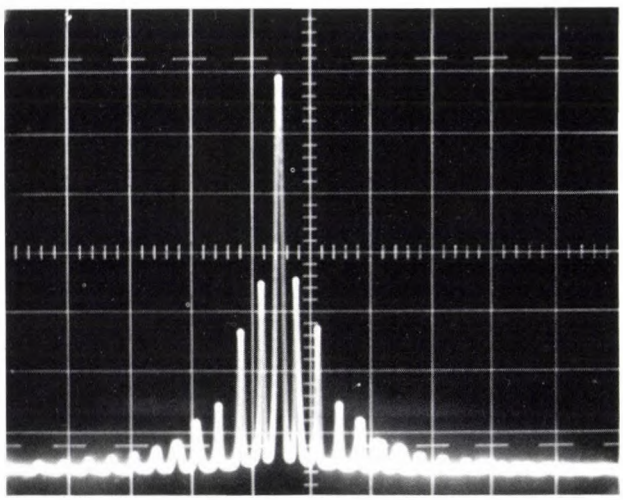


Fig. 8. Example of a spectrum of the light produced by the CQL10. The horizontal scale represents the wavelength; 0.85 nm per scale division. The peaks correspond to 'longitudinal modes of vibration' of the light in the laser cavity.

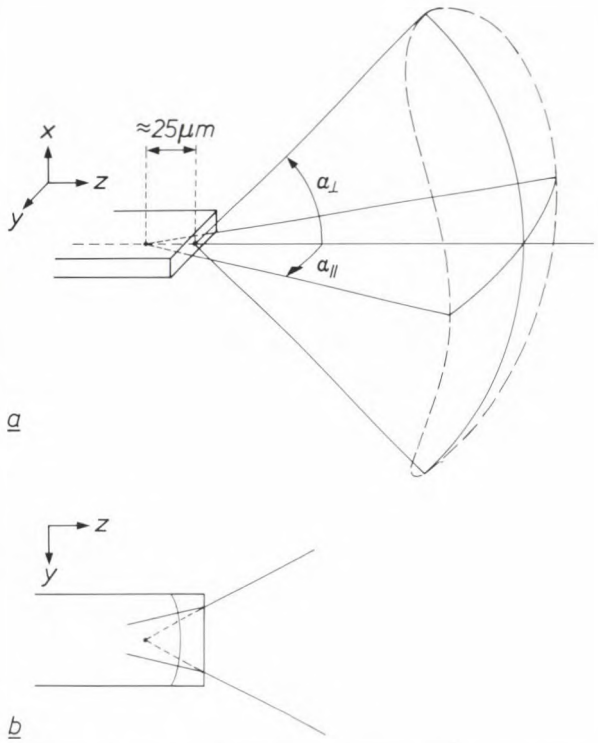


Fig. 10. The astigmatism of the laser light. *a*) The centre of curvature of a wavefront in the x-direction lies in the plane of the exit mirror and the centre of curvature in the y-direction is about 25 μm behind it. *b*) The astigmatism of the light arises inside the laser, where a wavefront is curved in the y-direction because of 'gain guiding'.

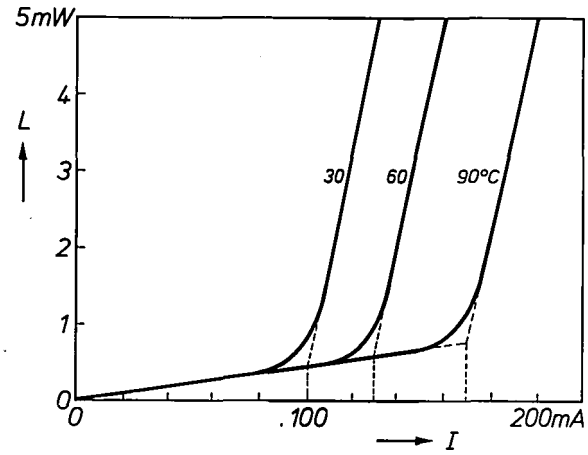


Fig. 11. Temperature dependence of the light/current characteristic. The threshold current increases as the temperature increases. In this example the increase is 30% per 30 K.

the wavefront of fig. 10*a* is produced. The astigmatism can be corrected by a cylindrical lens.

The light/current characteristic (fig. 3) is dependent on the temperature of the laser crystal (*fig. 11*). The threshold current increases with temperature; the variation is expressed empirically by

$$\frac{I_{th2}}{I_{th1}} = \exp \frac{T_2 - T_1}{T_0}.$$

In the temperature range of fig. 11 the effect must be attributed to the higher probability of electrons and holes escaping across the barriers ΔE_c and ΔE_v (fig. 2) at a higher temperature. The quantity T_0, which characterizes the temperature behaviour, depends therefore to a large extent on ΔE_c and ΔE_v. T_0 is at least 115 K for the CQL10. This means that the threshold current increases by no more than 30% for a temperature rise of 30 K.

To limit the threshold current the temperature of the active region must be kept as low as possible. For this reason the laser crystal has been soldered with the upper layer on the assembly block; the active layer is then as close as possible to the block. The thermal resistance between the active region and an external heat sink is then about 25 K/W; about 5 K/W of this is due to the block, as we saw earlier. If for one reason or another the active region is not cooled adequately the laser may go into 'thermal runaway'; this is explained in *fig. 12*.

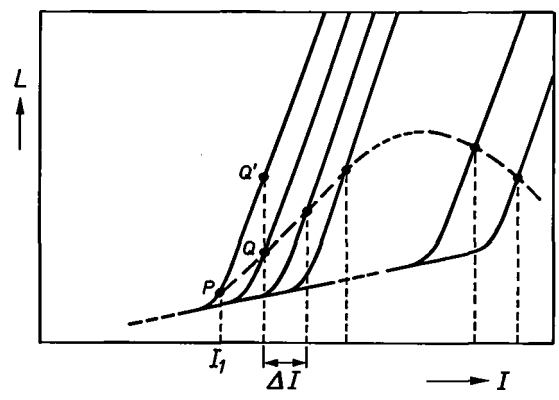


Fig. 12. 'Thermal runaway' of a laser. If the active region is insufficiently cooled — because the thermal resistance between this region and block is too high, too much heat is developed in the active region or the block is insufficiently cooled — the temperature of the active region rises and its characteristic curve (the solid lines) shifts to the right when the current is increased. Starting from a current I_1 a current increase ΔI does not take the operating point to Q' as it would have done if the temperature in the active region had remained the same, but to Q. Initially, therefore, the characteristic is less steep. As the current is increased, the shift for each step ΔI becomes increasingly larger so that in the end even *less* light leaves the laser and it goes into thermal runaway.

In measuring the thermal resistance we make use of the temperature dependence of the refractive index. The position of the spectrum is determined for one temperature (T_1) of the block and one current, then the temperature of the block is reduced to $T_1 - \Delta T$ and the current is adjusted so that the spectrum comes back to the same place. The active region then has the same temperature as before. The thermal resistance is $\Delta T/\Delta W$, where ΔW is the increase in the electrical power supplied (which is almost completely converted into heat).

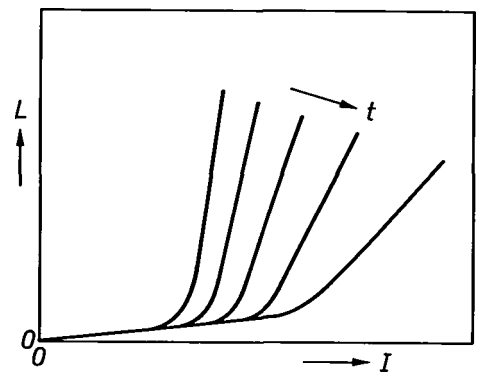


Fig. 13. Deterioration of the laser. With the passage of time the characteristic curve becomes less steep and the threshold current becomes higher. If the light level is too high or too much heat is generated in the laser this deterioration takes the form of instantaneous damage, but when suitable precautions are taken the effect is not measurable even after several thousands of hours of operation.

The laser gradually deteriorates during use; the threshold current increases and the light/current characteristic above the threshold becomes less steep (*fig. 13*). The rate of deterioration, however, depends strongly on the temperature; a rise of 30°, for example, reduces the life by a factor of about 15. If

care is taken to keep the temperature under control, the laser will have a long life; life tests have shown that at a light level of 5 mW and an ambient temperature of 30 °C the quality of many lasers has hardly changed after 10 000 hours of operation.

There are various reasons for the deterioration in performance. In the first place, crystal defects that have formed outside the active region during growth or during assembly may grow into this region. Defects may also change position under the action of the injection current and so penetrate into the active region. (It may also happen that defects leave the active region, as is evident from the improvement in the light/current characteristic that is sometimes observed during 'burning in'.) Finally, as noted earlier, the mirrors are sensitive elements since they contain many crystallographic defects because they are cleavage planes. These defects induce non-radiative heat-producing recombination processes that cause chemical changes in the material, and they can also induce a kind of punch-through effect at very high light intensities. These various effects rarely occur at room temperature but are activated by the light and heat developed in the operating laser.

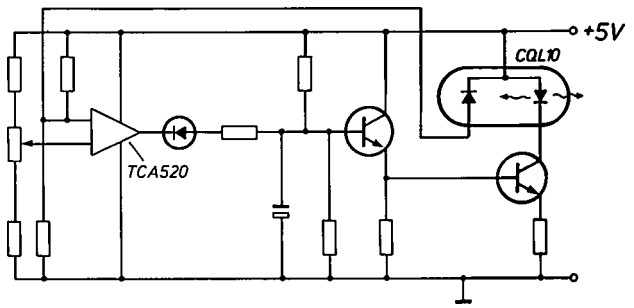


Fig. 14. Circuit for stabilizing the light flux from the laser. In the diagram of the laser (*CQL10*), the diode symbol on the right represents the laser crystal, the one on the left represents the photodiode. The circuit contains feedback; the photodiode records the light flux from the laser and controls the supply current so as to correct for variations in the light flux.

Electrically, the CQL10 is a diode with a knee voltage of about 1.6 V and a series resistance of about 2Ω. To obtain a current in the forward direction, it is preferable, because of the steepness of the *I-V* characteristic at the operating point, for such a diode to be supplied from a current generator. However, if the laser is directly connected to a current generator, the light level will tend to fluctuate because the *L-I* characteristic is also very steep and may vary as a result of temperature fluctuations or ageing. The photodiode behind the laser crystal in the CQL10 can be used for stabilization of the light level, which is required in many applications. *Fig. 14* gives an example of a power-supply circuit that will keep the light level constant with the aid of the photodiode.

When in use, the laser must be carefully protected from current surges that produce excessively high optical fluxes; in such cases deterioration is so fast that it is more accurate to speak of instantaneous damage. Such damage may even be caused by relatively small current peaks of very short duration (nanoseconds) because the characteristic curve is so steep, and the laser is also so fast that it can follow variations at frequencies in the gigahertz region. It is therefore of primary importance to protect the laser from voltage peaks, e.g. from the mains supply. Precautions should also be taken when switching on or off, since this can often cause large variations in current; it is recommended that the power supply should be turned down and the laser short-circuited before switching on or off.

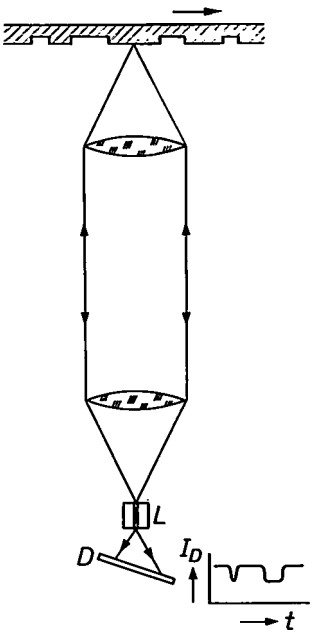


Fig. 15. Diagram illustrating detection with feedback. The light that is reflected by the disc is fed back to the exit mirror of the laser crystal (*L*). The laser reacts to this by an increase in the emitted power. Fluctuations in the reflected light are therefore reproduced in the current I_D in the detector *D* behind the laser crystal.

Detection of the information signal by feedback

Finally, we shall discuss the possibility of detecting the information in the reflected light by means of the laser itself. In this case the reflected light is not coupled out (see fig. 1), but fed back to the exit mirror of the laser (*fig. 15*). The laser crystal then reacts to variations in the light reflected by the disc with variations in the emitted power; these are detected by the photodiode [9]. (When a stabilizing circuit like the

one shown in fig. 14 is used, the variations should of course be fast compared with the response time of this circuit.)

On closer examination of the 'feedback effect' it appears that the laser reacts to feedback with a shift of the light/current characteristic towards lower current values (*fig. 16a*); for a given current intensity, therefore, more power is obtained from the laser with feedback than without. In the first instance this is quite understandable: because radiation is fed back, the laser has effectively smaller radiation losses (a higher Q) and — for the same current — it therefore starts to oscillate at a higher level. The feedback effect (fig. 16*b*). F is also highly dependent on the external optical system, of course.

We use the arrangement of *fig. 17a* to test laser crystals for sensitivity to feedback. When the mirror S is located in the image of the laser mirror ($z = 0$) about 2% of the light is fed back. This is reduced to zero by even a very small shift in S. The photodiode signal as

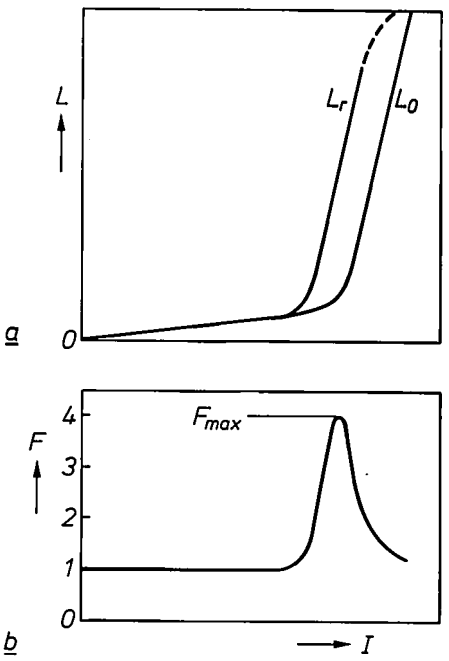


Fig. 16. *a)* The L-I characteristic without feedback (L_0) and with feedback (L_r). In the main the effect is to decrease the threshold current: *b)* 'Feedback factor' $F (= L_r/L_0)$ as a function of I. It can be seen from (*a*) that F must have a maximum when the current is close to I_{th}.

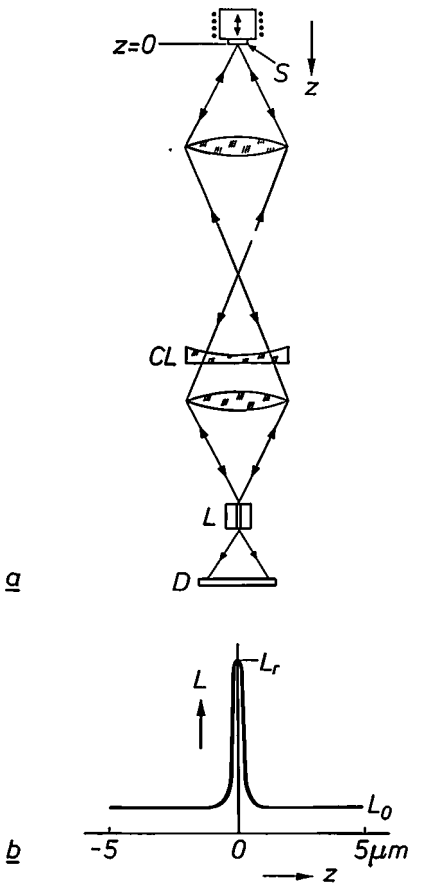


Fig. 17. *a)* Arrangement for measuring the feedback factor. L laser, D detector. The mirror S is placed on a vibrating loudspeaker cone. CL cylindrical lens. *b)* The radiated power as a function of the position z of the mirror.

is really a very complicated effect: the laser cavity is coupled to an 'external cavity' (laser mirror, optical system and reflector) and this coupling results in a fine structure for every laser mode [10]. We shall not go into these complications in detail, but merely point out that the reaction of the laser here is relatively simple because the coherence length is much smaller than the optical pathlength.

In detection we are concerned with the 'feedback factor' F, i.e. the ratio L_r/L_0 of the optical power with feedback to that without feedback. It follows from fig. 16*a* that F plotted as a function of the current will have a maximum (F_{max}) in the vicinity of I_{th} a function of z thus provides the feedback factor directly for this arrangement (fig. 17*b*). The mirror is attached to a vibrating loudspeaker cone and the curve of fig. 17*b* can be read off an oscilloscope. With this arrangement we find values for F_{max} that vary from 5 to 20 depending on the dimensions and composition of the wafer from which the laser crystal originates.

A device for reading out information stored in

[9] See for example the article of note [1].

[10] Y. Mitsuhashi, T. Morikawa, K. Sakurai, A. Seko and J. Shimada, Optics Comm. **17**, 95, 1976.
C. H. F. Velzel and R. P. Brouwer, IEEE J. **QE-15**, 782, 1979.
C. H. F. Velzel, Ned. T. Natuurk. **A45**, 54, 1979.

tracks of pits as on a 'VLP' disc is a very complex unit. Whether the detection method outlined here will be used for this purpose depends on a complex of factors that we shall not discuss here. In the CQL10, however, this possibility has been taken into account in the sense that the photodiode that detects the laser power is fast enough to follow the variations that represent the information flow.

Summary. The CQL10 is a semiconductor laser for read-out of the information on a 'VLP' disc, a 'Compact Disc', a 'DOR' disc or similar devices. It is about 1 cm in size and operates from a voltage of 2 to 3 V. The laser crystal is a substrate of GaAs with epitaxial layers of $Al_xGa_{1-x}As$. With $x = 0.16$ for the 'active' layer the wavelength obtained is short enough (780 nm) for the intended application. The value of x for the 'cladding' layers is 0.46. Proton implantation is used to limit the flow of current and the laser activity to a narrow stripe 5 μm wide. The end mirrors of the laser cavity are formed by cleavage planes. The crystal, in good thermal contact with a heat sink, is mounted in an encapsulation with a window and connecting pins. The threshold current for laser operation is about 120 mA at 30 °C and has not increased by more than 30% at 60 °C. The light/current characteristic is very steep above the threshold. The spectrum consists of lines at a spacing of about 0.3 nm. The beam is strongly divergent and is slightly astigmatic. Laser deterioration is not usually noticeable even after 10 000 hours of correct operation, but care must be taken to avoid voltage peaks so as to protect the mirrors. A photodiode in the encapsulation can be used for electronic stabilization of the light flux and for detection of the information signal by feedback to the laser.

Measuring the magnetic properties of a microscopic particle

In studying the magnetic properties of recording tape it is of interest to know the properties of the individual magnetic particles. *Fig. 1* shows an electron-microscope photograph of the surface of a magnetic tape, in which the magnetic particles are visible. The particles whose properties are to be measured are detached from the tape. Their lengths typically lie between 0.5 and 1 μm; the length/thickness ratio is about 7.

Among the methods developed over the years to investigate small grains of magnetic material [1] there is only one suitable for the particles with which we are concerned here [2], but it requires an extremely sensitive torque magnetometer, which is very fragile.

The new method described is based on a determination of the rate at which a magnetized particle rotates in a suspension when subjected to a weak opposing external field. This results in a 'remanence curve', examples of which are shown in *fig. 2* for a γ-Fe_2O_3 particle and a CrO_2 particle.

To make such a measurement possible the particles are dispersed in a very viscous polymer solution, in which they are separated so far apart that interaction between them is negligible. This suspension is drawn into a flat capillary tube, making it possible to observe the particles with an optical microscope. By means of the microscope and a television camera a particle selected for observation is magnified 10 000 times, so that it is displayed with a length of about 10 mm on a television monitor (fig. 1, *inset*).

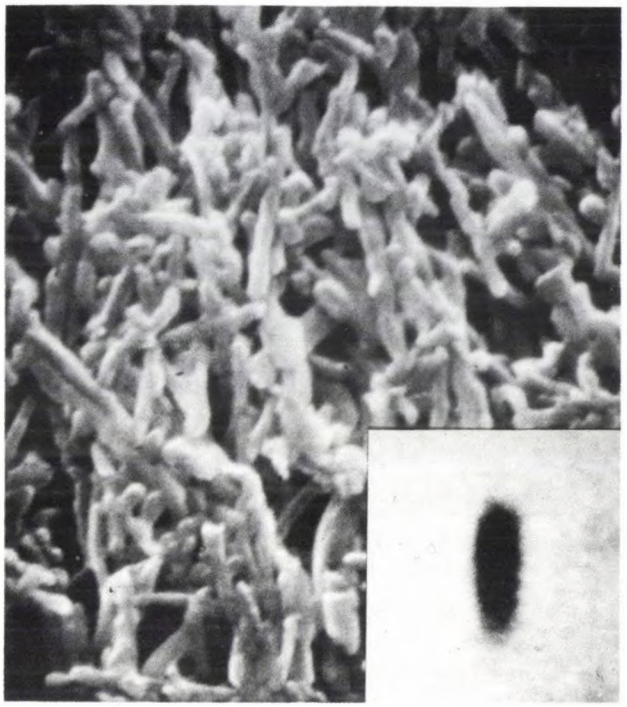


Fig. 1. Magnification, made with a scanning electron microscope, of iron-oxide particles (γ-Fe_2O_3) in the surface of an etched magnetic tape. *Inset:* An individual particle, displayed on a monitor screen after measurement by means of an optical microscope and a television monitor. The actual length is about 1 μm, the estimated thickness 0.15 μm. (Owing to the limitations of the optical system the width of the particle in the display is greater than it is in reality.)

In more detail, the curves are determined as follows. The particle is subjected continuously to a steady field H_{bias} of 640 A/m, which is applied by

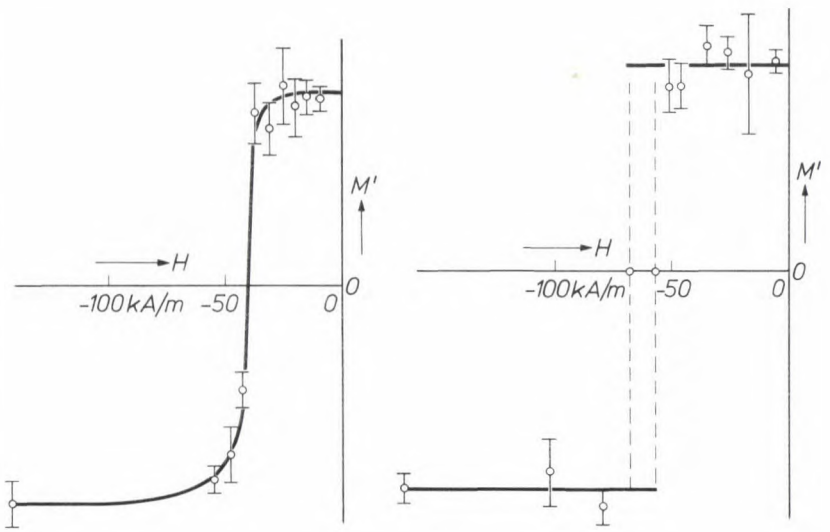


Fig. 2. Two measured remanence curves, with the magnetizing field H along the abscissa and the remanent magnetization M' along the ordinate, in arbitrary units, for a γ-Fe_2O_3 particle (*left*) and a CrO_2 particle (*right*).

[1] H. Zijlstra, Philips tech. Rev. **31**, 40, 1970.
[2] S. P. Yu and A. H. Morrish, Torsion balance for a single microscopic magnetic particle, Rev. sci. Instr. **27**, 9-11, 1956.
[3] For an evaluation of the measurements, see: J. E. Knowles, Measurements on single magnetic particles, IEEE Trans. **MAG-14**, 858-860, 1978, and J. E. Knowles, Magnetic measurements on single acicular particles of γFe_2O_3, IEEE Trans. **MAG-16**, 62-67, 1980.

means of a coil consisting of a few turns wound around the capillary. Before each measurement the particle thus aligned in the direction of H_{bias} is first fully magnetized by means of a field pulse of 120 kA/m, applied for 20 μs through the same coil. In *fig. 3*, which shows a 'normal' magnetization loop, the particle has acquired the value of M_{max} after this field pulse. Next, a smaller pulse H_A is applied in the same manner. The particle then first acquires a magnetization of M_A (point A) and it retains a remanent magnetization of M_A' after the pulse.

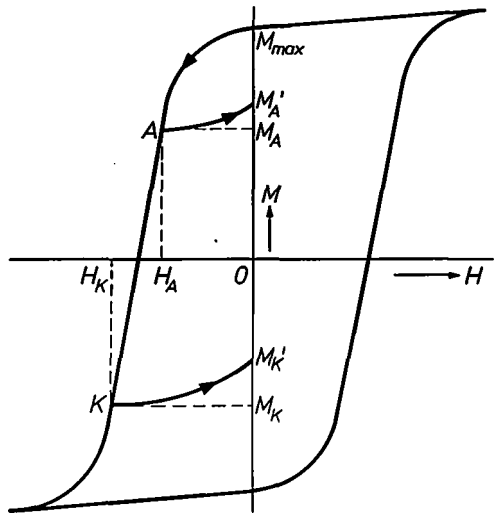


Fig. 3. Schematic representation of the 'normal' hysteresis loop of ferromagnetic material — magnetization M plotted against field-strength H. In the measurement on single magnetic particles of magnetic tape the particles are first fully magnetized with a strong field pulse ($M = M_{max}$), and are then subjected to a second field pulse H_A. While the second field is present the magnetization acquires the value M_A; upon its removal the value becomes M_A'. The latter value is measured. If the field pulse has a strength H_K greater than the coercive force, the resulting remanent magnetization M_K' is in the opposite direction.

To measure the value of M_A' the field H_{bias} is now abruptly reversed. Because of friction with the surrounding viscous fluid, the particle rotates relatively slowly, making it possible to determine the time it takes to rotate through an angle of say 90°. It is assumed that this time is inversely proportional to the remanent magnetization M_A' of the particle, which is subsequently recorded in arbitrary units.

If the second field pulse was so strong as to make the remanent magnetization negative, giving it for example a value M_K' (fig. 3), then the particle spontaneously rotates in the field $+H_{bias}$.

The measured points in fig. 2 were obtained by determining the rotation rate of the particle in a given magnetic state five times. The short lines indicate the spread in the results [3].

J. E. Knowles

J. E. Knowles, B.Sc., is with Philips Research Laboratories (PRL), Redhill, Surrey, England.

DKDP crystals for use in the TITUS tube

C. Belouet

Single crystals of deuterated potassium dihydrogen phosphate (DKDP) give a strong electro-optical effect at a temperature close to their Curie point. This property is the basis for the operation of the TITUS tube developed at LEP (Laboratoires d'Electronique et de Physique Appliquée), in which a single-crystal DKDP target acts as a solid-state light modulator for the large-screen projection of television pictures. Earlier articles in this journal have dealt with the operation of the tube and modifications to it, but no attention has been paid to the relation between the quality of the DKDP crystal target and that of the projected pictures. The article below shows how crystal defects can have a considerable effect on the electro-optical properties of the target, with a deterioration in picture quality. Research on the origin of these defects has produced much useful information about the most appropriate conditions for growing DKDP crystals for use in the TITUS tube.

Introduction

DKDP is an abbreviated representation for deuterated potassium dihydrogen phosphate, $KH_{2(1-x)}D_{2x}PO_4$. A tetragonal DKDP crystal is ferroelectric and gives a strong electro-optical effect near the Curie point in the paraelectric phase. An electric field applied along the crystallographic [001] axis induces a birefringence for light propagating in this direction. The induced birefringence is proportional to the electric polarization, i.e. to the product of the dielectric constant and the electric field-strength.

This effect is utilized in the TITUS tube, an optical relay tube for the large-screen projection of television pictures [1]. The principle of operation of the TITUS tube is shown schematically in *fig. 1*. Scanning with an electron beam of constant intensity, which is modulated by a video signal, produces a pattern of electric charges on the back of a crystal plate of DKDP (the target). The birefringence induced by the electric charges results in a one-to-one modulation of an external plane-polarized light beam, which enters the DKDP target through the front face. After being reflected at the rear face and passing through the target again, the modulated light beam goes through a second polarizer (the analyser) and is then projected on to a large screen. In this way an optical reproduction of the electrical video signal is obtained via the latent image in the DKDP target.

To obtain good reproduction there must be no stresses in the crystal that will produce false contrasts (photoelastic contrasts) in the picture, and the decay of the stored images must be slow and uniform. This implies both the absence of deformations in the crystal and a low, homogeneous electrical conductivity. In addition, the crystal target should be fairly large; a typical target surface area is about 4×3 cm.

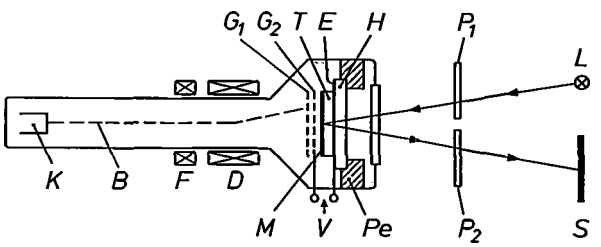


Fig. 1. Diagram illustrating the operation of the TITUS tube. The back of a single-crystal target *T* is scanned by an electron beam *B* of constant intensity from a cathode *K*. The beam is focused and deflected by coils *F* and *D*. A short distance away from the target are two collector grids G_1 and G_2. The video signal *V* is applied to the conducting transparent electrode *E*. The birefringence induced in the target modulates a light beam incident on the front face of the target. The light beam, originating from light source *L*, is polarized by polarizer P_1, and is reflected from the rear of the target by a thin-film mirror *M*, so that the beam passes through the target again. It then passes through a second polarizer P_2, which is crossed with respect to P_1, and is finally projected on a screen *S*. *H* fluorite holder. *Pe* Peltier cooler.

Dr C. Belouet is with Laboratoires d'Electronique et de Physique Appliquée (LEP), Limeil-Brévannes (Val-de-Marne), France.

An important feature of the TITUS tube is that the target is cooled to just above the Curie point of the crystal. This produces a stronger electro-optical effect and at the same time a higher electrical resistivity, so that the images can be stored for much longer. The Curie point of DKDP crystals rises sharply with the deuterium content x, from -152 °C for $x = 0$ to -55 °C for $x = 1$. Thus, if x is given a value close to 1, the crystal needs far less cooling.

DKDP crystals are grown from aqueous solutions. *Fig. 2* shows schematically the appearance of a DKDP crystal in the ideal case. The crystal is tetragonal and

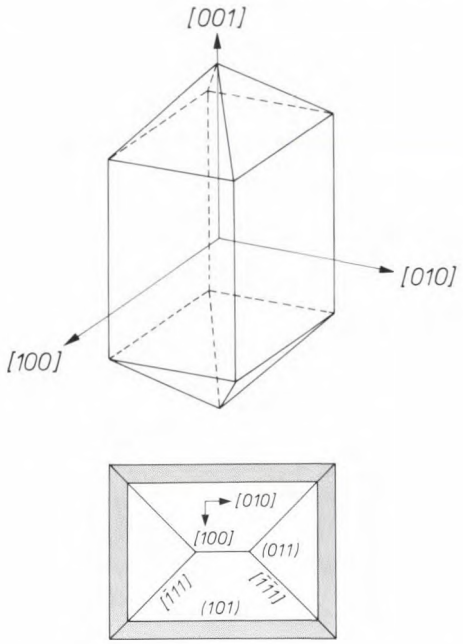


Fig. 2. *Above:* ideal habit of a tetragonal DKDP crystal. *Below:* diagrammatic cross-section in the (001) plane with the boundaries between the different growth sectors. There are four pyramidal sectors and four prismatic sectors (*grey shading*).

consists of a number of flat regions (facets) that together form a four-sided prism, with a tetragonal pyramid at the top and at the bottom. During growth, sectors of two different kinds can be distinguished: prismatic sectors, corresponding to growth in the [100] and [010] directions, and pyramidal sectors, corresponding to growth in the [001] direction, which takes place along the four pyramidal facets (101), ($\bar{1}$01), (011) and (0$\bar{1}$1). In practice, deviations from the ideal habit may occur, and these result in a tapering of the crystals along the [001] axis; see *fig. 3*.

Under usual growth conditions, pyramidal growth is predominant, so that the crystals become elongated in the [001] direction, at a typical rate of about 0.3 mm a day. The results of the continuous growth of four neighbouring and interacting pyramidal facets can therefore be seen in each DKDP target. The crystal defects thus produced make it difficult to meet the TITUS requirements.

The investigations described in this article were aimed at finding out more about the origin of these defects and about their effect in generating flaws in displayed images. The results enabled us to develop a process for growing DKDP crystals that give improved performance.

In this article the growth of DKDP crystals and the determination of the properties of the grown crystals in relation to those of the growth solutions will first be

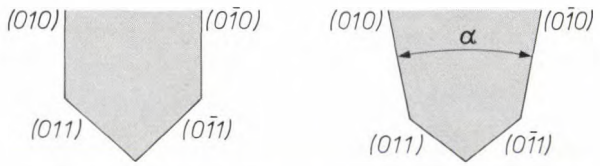


Fig. 3. Diagrammatic (100) section through an ideal DKDP crystal (*left*) and through a DKDP crystal characterized by some tapering, with an angle α (*right*).

dealt with in more detail. The effect of crystal defects on the electrical conductivity will be discussed next, and it will be shown how these crystal defects occur and how they can be suppressed by taking proper precautions.

The growth of DKDP crystals

There are various methods for growing DKDP crystals from aqueous solutions. They differ from each other in the way in which the supersaturation necessary for the growth is established.

Processes where growth occurs at a constant temperature are based on the considerable influence of the acidity (pH) on the DKDP solubility, the sharp increase of the solubility with temperature, or the high solubility in a wide temperature range; see *fig. 4*. The pH dependence can be exploited in a process in which the 'driving force' comes from the gradual addition of concentrated potassium-hydroxide solution to a saturated DKDP solution with a low pH value [2]. The strong temperature-dependence of the solubility can be exploited by causing a transport to take place from a vessel with a higher temperature to

[1] 'TITUS' is an acronym for 'Tube image à transparence variable spatio-temporelle'. More details are given in G. Marie, Philips Res. Repts. **22**, 110, 1967, G. Marie, Philips tech. Rev. **30**, 292, 1969, and J. Donjon and G. Marie, Philips tech. Rev. **34**, 129, 1974.

[2] I. F. Nicolau, M. Ittu and R. Dabu, J. Crystal Growth **13/14**, 462, 1972.

a 'growth jar' with a lower temperature [3]. Modified versions using a single jar divided into two parts at different temperatures have also been proposed [4]. The high solubility of DKDP combined with the easy evaporation of the solvent has resulted in the widely used slow-evaporation technique, in which a controlled stream of a dry gas is passed over a saturated solution.

Besides the methods that utilize a constant growth temperature, there are methods in which the supersaturation is maintained by reducing the temperature at a monitored rate [5].

Although steady-state processes, i.e. those using a constant temperature, seem at first sight to be more attractive, they have the great disadvantage that their reliability over extended periods of growth often leaves something to be desired. This arises from the need to control an extra variable in addition to the temperature, and from the occurrence of unforeseen nucleations ('volunteer growth'), which compete with single-crystal growth and make the process difficult to control (they result in an uncontrolled 'driving force'). The temperature-decrease method was therefore chosen in order to obtain large single crystals (boules) of good quality.

The vessel used for most of the experiments is shown schematically in *fig. 5*. The growth chamber, which has a capacity of five litres, is immersed in an outer bath through which water circulates at a controlled temperature (maximum variation 0.02 °C). The water, which comes from a thermostat set to the appropriate temperature, also flows through a polymethyl methacrylate lid on the growth chamber. The main features of this equipment are that the [001] direction of the crystal grown on the seed points downwards, and that the movement of the solution past the crystal faces is primarily controlled by the steady rotation of the crystal about its holder axis, while the solution is homogenized by three propellers rotating in the opposite direction to the crystal.

Other vessels are usually used for growing DKDP crystals on an industrial scale. The accuracy of the temperature control in the inner bath of such vessels is about 0.01 °C [6] and the temperature deviations in the growth solution are smoothed out by the larger capacity of the growth chamber (about 65 litres). A difference from the experimental equipment is that the seeds are mounted vertically on a 'seed tree', so that growth in the [001] direction takes place horizontally. The solution is stirred by the rotation of the seed tree, with a simple d.c. motor so that the direction of rotation can be reversed.

The final quality of the grown crystal does not only depend on the performance of the growth equipment; it is equally dependent on the control of the cleanliness of the parts exposed to the growth solution, the preparation of the solution and the quality of the seed crystal.

Cleanliness of the growth chamber prevents direct contamination of the solution and reduces the risk of undesirable nucleations. The growth chamber is therefore made air-tight and all the components (jar, lid, seed, holder, etc.) are soaked in a detergent,

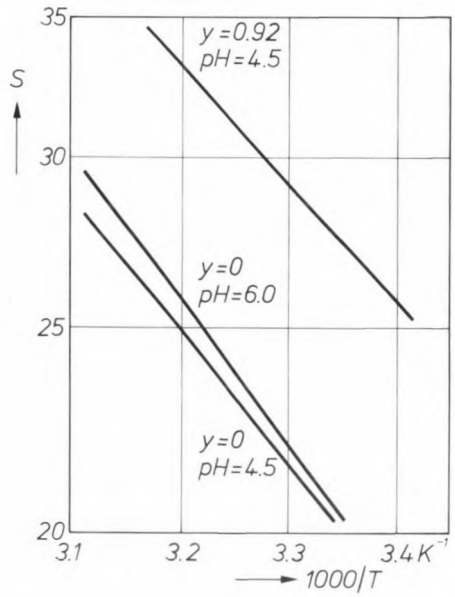


Fig. 4. Solubility S of DKDP expressed in grams per 100 g of aqueous solution, plotted on a logarithmic scale as a function of the reciprocal of the absolute temperature, for a deuterium content $y = 0$ at two different pH values (4.5 and 6.0), and for $y = 0.92$ at a pH of 4.5.

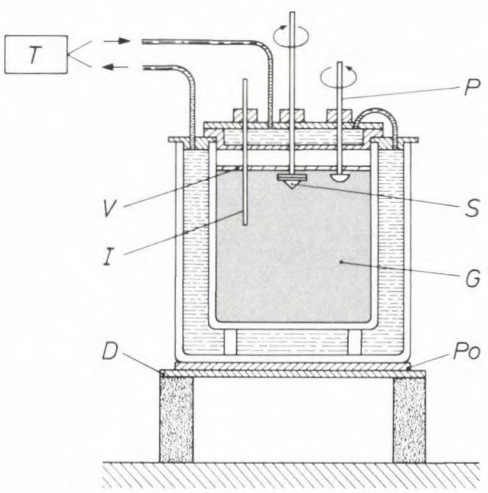


Fig. 5. Diagram of a vessel for growing DKDP crystals. On cooling of a supersaturated DKDP growth solution G, crystal growth takes place on the rotating seed crystal S. The growth solution is introduced into the growth chamber through the inlet I. The solution is stirred mechanically by three equidistant propellers P, and can be covered with a layer of Vaseline oil V as an insulation. To keep the growth solution at the right temperature, water flows via thermostat T through the growth chamber lid and the surrounding bath. D Duralumin. Po polystyrene.

rinsed in hot water and exposed to a fresh and concentrated aqueous solution of ammonium fluoride. They are then washed in filtered water of high resistivity and dried out at room temperature in a stream of dry and filtered nitrogen.

The solutions are preferably prepared by direct synthesis from heavy water, phosphorus pentoxide (P_2O_5) and potassium carbonate (K_2CO_3). The P_2O_5 reacts with heavy water to produce a solution of deuterated phosphoric acid. After being refluxed and filtered, this solution is allowed to react slowly with a filtered K_2CO_3 solution in a sealed vessel through which a dry inert gas (nitrogen) flows. After completion of the reaction, the DKDP solution is heated to about 50 °C and stirred vigorously to remove any dissolved carbon dioxide. For this operation a polypropylene vessel is preferable to a Pyrex one, since erosion of the Pyrex walls at this temperature contaminates the solution. After filtration, the solution is ready for use.

The seeds used are (001) plates cut from non-tapered crystals of good quality, with a deuterium content close to the value for equilibrium with the solution. The seed plates are polished mechanically and chemically on PTFE (Teflon) discs with a mixture of acetic acid, ethanol and water, and are then mounted in the empty growth chamber. There they are first subjected to additional etching by pouring in the growth solution at a temperature slightly above its saturation point. Then the growth process is started by a programmed decrease in temperature and stirring the solution. In this procedure, nucleation on the seeds takes place at a very low supersaturation, corresponding to a temperature decrease of about 0.01 °C.

The choice of the initial temperature and the *p*H value of the growth solution is restricted by the requirement that the DKDP crystals must have a high deuterium content. Crystals grown from a heavily deuterated solution at a relatively high initial temperature are found to have cracks. This is related to a narrowing of the existence domain of the tetragonal phase of DKDP with increasing values of the deuterium content *y*. On the high-temperature side of the solubility curve DKDP forms a monoclinic variant, as shown in *fig. 6* for neutral solutions [7]. At high temperatures and *y*-values the tetragonal phase is therefore not stable with respect to the mother liquor. It is observed that the tetragonal-to-monoclinic transition shifts to higher temperatures when the *p*H value of the growth solution is lowered. Thus, stable tetragonal boules can be grown from a solution with $y = 0.994$, up to an initial temperature of about 40 °C, if the *p*H value of the growth solution is lowered from the neutral value (4.5) to 2.2. This finding restricts the growth to the low *p*H range between 2.0 and 2.6, and the starting temperature should preferably be in the neighbourhood of 30 °C.

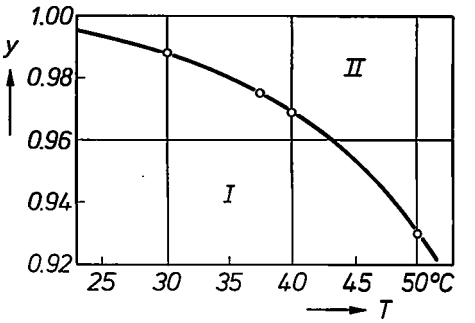


Fig. 6. Existence domains of the tetragonal phase *I* and the monoclinic phase *II* of DKDP, as a function of the temperature *T* and the deuterium content *y* in a neutral growth solution (*p*H = 4.5). As the deuterium content increases, the existence domain of the tetragonal phase becomes narrower.

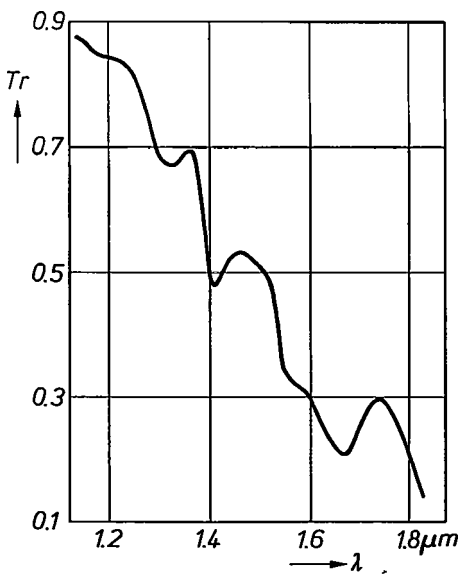


Fig. 7. Infrared transmission *Tr*, measured at room temperature, for a heavily deuterated DKDP solution, as a function of the wavelength λ. The spectrum has a sharp absorption band at about 1.4 μm.

[3] J. J. P. Valeton, Ber. math.-phys. Klasse Kön. Sächs. Ges. Wiss. Leipzig **67**, 1, 1915.
V. F. Parvov, Sov. Phys. Crystallogr. **12**, 324, 1967.
M. I. Kozlovskii, A. V. Komolobai, I. I. Melentyev and M. F. Burchakov, in: N. N. Sheftal' (ed.), Growth of crystals Vol. 6A, p. 7; Consultants Bureau, New York 1968.
[4] I. F. Nicolau, Kristall und Technik **9**, 1331, 1974.
B. M. Bartlett, J. sci. Instr. **38**, 54, 1961.
[5] R. W. Moore, J. Amer. Chem. Soc. **41**, 1060, 1919.
J. L. Torgesen, A. T. Horton and C. P. Saylor, J. Res. Nat. Bur. Stand. **67C**, 25, 1963.
C. Belouet, Acta Electronica **16**, 339, 1973.
[6] G. M. Loiacono, Acta Electronica **18**, 241, 1975.
[7] C. Belouet, M. Monnier and R. Crouzier, J. Crystal Growth **30**, 151, 1975.

Determination of the properties of crystals and growth solutions

The deuterium contents x and y in DKDP crystals and growth solutions can be determined with the aid of infrared transmission spectra [7]. *Fig. 7* shows an example of a transmission spectrum measured at room temperature. *Fig. 8* gives the transmission at 1.43 μm as a function of y for heavily deuterated DKDP solutions ($y \geq 0.98$). The transmission is found to increase exponentially with the value of y. The hydrogen content $(1 - x)$ in the crystals was determined directly from the absorption at 1.60 μm of blocks 2 cm thick. The value of $(1 - y)$ was also determined indirectly by measuring the transmission at 1.43 μm of a mixture of H_2O and D_2O obtained by rapid distillation of the growth solution. It was assumed that the hydrogen/deuterium ratio of the vapour did not differ from that of the solution [8]. The experimental error was found to be less than 0.3%.

The concentrations of impurities known to affect the electrical conductivity and the morphology of the crystals were determined both in the growth solutions and in the crystals by means of a neutron-activation technique [9]. The distributions of two impurities, iron and calcium, were determined separately by autoradiographic analyses on crystals grown from solutions doped with ^{59}Fe and ^{45}Ca [10].

Spatially extended defects in DKDP crystals were made visible by means of X-ray topography, usually by Lang's method [11].

The photoelastic effects were observed directly in polarized light or in TITUS pictures.

A handy method of displaying heterogeneities in the electrical conductivity is the 'electrode-decoration technique' (EDT) [12]. This technique is based on two particular properties of DKDP: its ionic conductivity is very sensitive to impurity traces, and, because of this conductivity, it undergoes electrolysis resulting in the release of gases at the electrodes [13]. When a direct voltage is applied between two mirror-like metallic electrodes evaporated on two opposite faces of a wafer of DKDP, a local enhancement of the electrical conductivity is associated with greater bubble formation from these gases at the electrodes; see *fig. 9*. This results in a roughening of the surfaces, which can be observed visually.

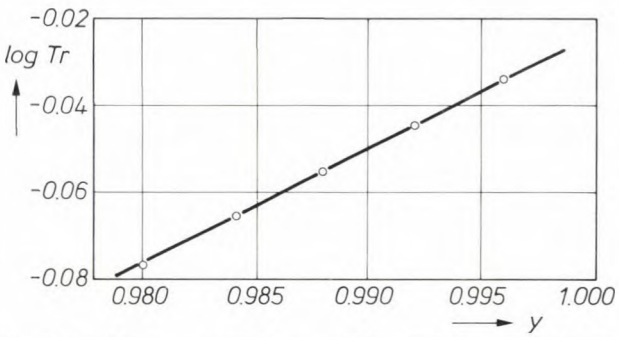


Fig. 8. Calibration curve that can be used for determining the deuterium content y of the DKDP growth solution. The logarithm of the infrared transmission Tr at 1.43 μm increases linearly with y.

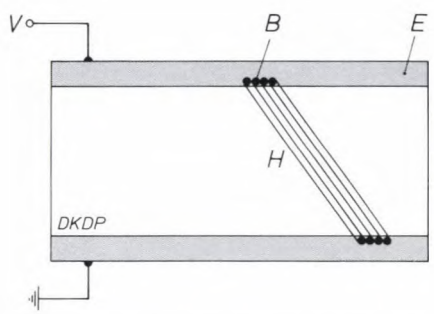


Fig. 9. Illustrating the electrode-decoration technique (EDT) for observing heterogeneities in the electrical conductivity. A (001) slice of the DKDP crystal under investigation is coated on opposite faces with two mirror-like metallic electrodes E. When a direct voltage is applied, a region H with a relatively high electrical conductivity gives rise to the formation of 'bubbles' B at the contact with each of the two electrodes.

Conductivity heterogeneities

It has been found from comparisons of EDT photographs with X-ray topographs and from chemical and autoradiographic analyses that defects in DKDP crystals can give rise to marked local enhancements of the electrical conductivity. *Fig. 10* shows an EDT photograph and an X-ray topograph of the same part of a DKDP crystal. The prismatic sectors have a far higher conductivity than the pyramidal sectors, and there is a clear relation between the heterogeneities in the pyramidal sectors and the crystal defects as revealed in the X-ray topograph. The heterogeneities in the pyramidal sectors are either found in regions delineated by crystal defects (e.g. sector boundaries and planar faults) or they are associated with extended strains in the vicinity of the seed; see *fig. 11*. Direct measurements of the electrical conductivity on separate parts of such heterogeneous sections revealed variations that could amount to orders of magnitude.

The magnitude of a local increase of the conductivity depends on the nature and the concentration of the various impurities. When crystals are grown from a solution, the content of a particular impurity is indicated by the segregation coefficient k. This is defined as the ratio of the concentration of the impurity in a crystal (or in a part of it) to the concentration in the growth solution. A comparison of the segregation coefficients $k(011)$ for the (011) sector of

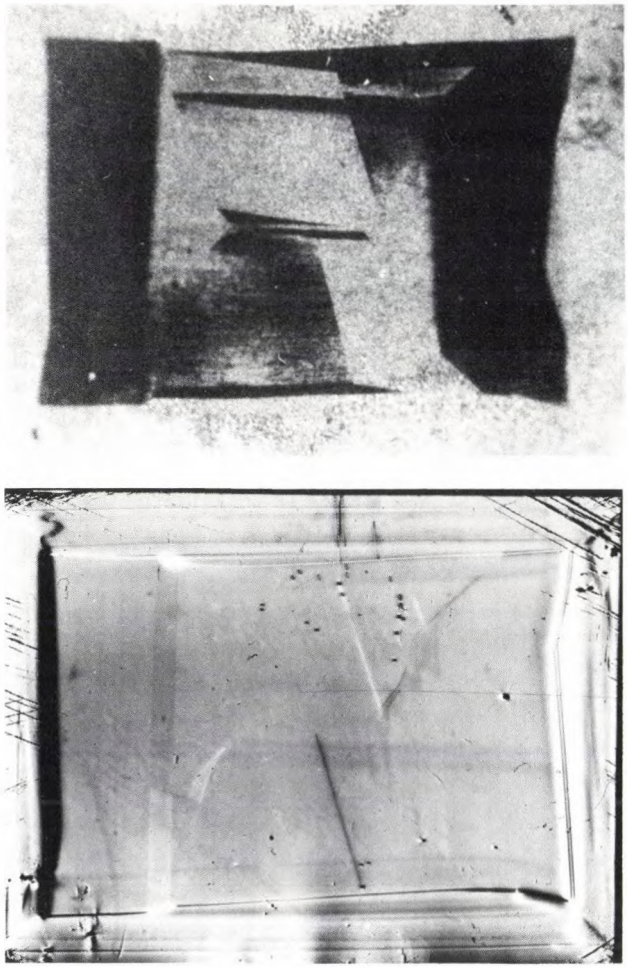


Fig. 10. EDT photograph (*above*) and X-ray topograph (*below*) obtained from a (001) slice of a DKDP crystal. The white areas in the EDT photograph indicate a relatively high electrical conductivity. It is clear that the locations of the boundaries between the regions of different electrical conductivity correspond to the locations of contrasts in the X-ray topographs. All X-ray topographs for this article were taken with MoKα_1 radiation; the reflections used are of the 200 type.

Table I. Segregation coefficient $k(011)$, defined as the ratio of the concentration of a particular element in the (011) sector to the concentration in the corresponding growth solution, for a number of elements in an unstrained and strained DKDP crystal. An unstrained crystal is taken to be a crystal whose X-ray topograph shows only weak contrasts in the sectors and at sector boundaries.

Element	$k(011)$	
	Unstrained crystal	Strained crystal
Cerium	0.017	4
Cobalt	0.022	0.65
Chromium	0.027	1.4
Caesium	6.0	0.64
Iron	0.05	4
Rubidium	0.04	0.18
Antimony	0.024	1.5
Strontium	0.13	0.73
Zinc	0.0125	0.28

a number of impurities reveals that the segregation coefficient for strained crystals is generally much greater than for non-strained crystals; see *Table I*. This indicates a close connection between crystal strains and the incorporation of impurities during growth [14]. The mechanism of this coupling was not elucidated in the present study. What has definitely been established, however, is that although the growth solution must have a high purity if crystals of low conductivity are to be obtained, it is equally important to suppress strains during the growth. An additional complication is found in the growth of heavily deuterated crystals, because the deuterium content and hence the crystal lattice parameters are sensitive to the growth rate owing to the high value of the hydrogen segregation coefficient [7,8,15].

Heterogeneities in the electrical conductivity of a DKDP crystal may give rise to undesirable contrasts in TITUS displays. These contrasts have a clear one-to-one correspondence with the heterogeneities in the EDT photograph of the same crystal; see *fig. 12*.

In the following section closer attention will be given to the origin of the crystal defects, and to finding the experimental conditions in which these defects and the resulting conductivity heterogeneities can be prevented.

The origin of crystal defects and their prevention

A particular feature of the growth of DKDP crystals is the occurrence of crystal imperfections inherent in the seeding process, in addition to the growth-induced defects. In the following a distinction will therefore be made between these two types of defect. Special attention will be paid to the interaction of defects that arise as a consequence of the overlapping of adjacent pyramidal sectors with different rates of growth.

Seed-induced defects

In the present state of the art, DKDP crystals of high quality are grown on (001) seed plates rather than on seeds that have the natural tetragonal habit or

[8] R. S. Momtaz and L. N. Rashkovich, Phys. Stat. sol. (a) **38**, 401, 1976.
L. A. Dmitrenko and V. V. Korolikhin, Sov. Phys. Crystallogr. **20**, 282, 1975.
[9] See for example M. L. Verheijke, Philips tech. Rev. **34**, 330, 1974.
[10] C. Belouet, M. Monnier and J. C. Verplanke, J. Crystal Growth **29**, 109, 1975.
[11] A. R. Lang, J. appl. Phys. **30**, 1748, 1959.
[12] C. Belouet and M. Monnier, Acta Electronica **18**, 143, 1975.
[13] M. Sharon and A. K. Kalia, J. solid State Chem. **21**, 171, 1977.
[14] C. Belouet and W. T. Stacy, J. Crystal Growth **44**, 315, 1978.
[15] G. M. Loiacono, J. F. Balascio and W. Osborne, Appl. Phys. Lett. **24**, 455, 1974.

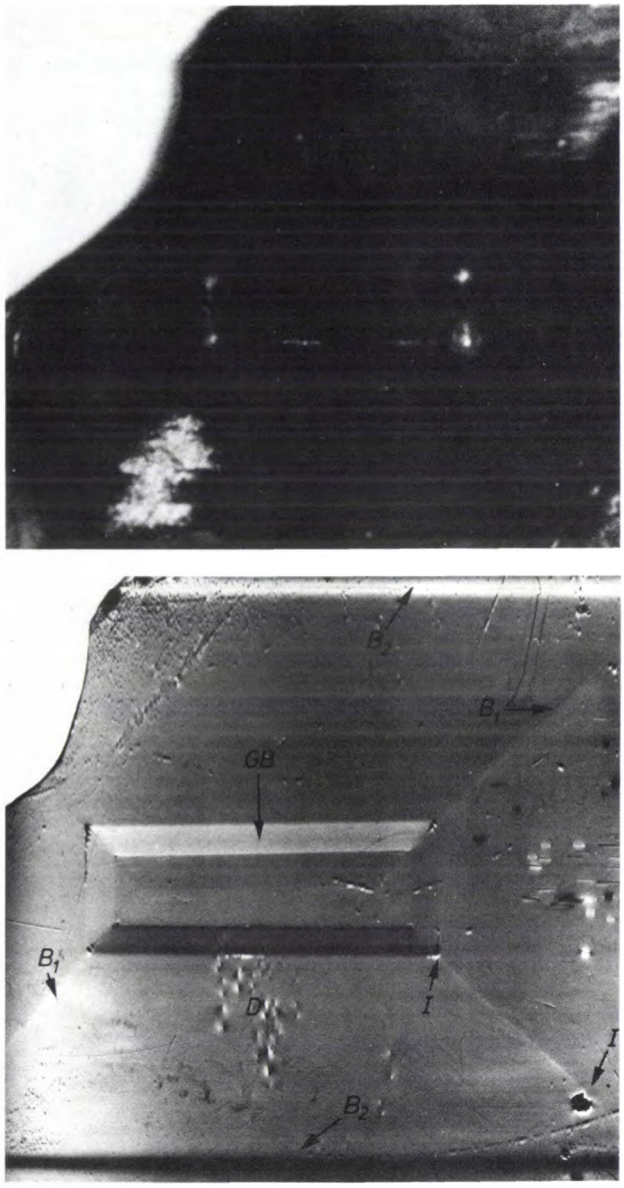


Fig. 11. EDT photograph (*above*) and X-ray topograph (*below*) of a (001) slice of DKDP sawn off near the end of the tetragonal cap (the final growth habit). A white area in the EDT photograph indicates a region of relatively high electrical conductivity. Increased conductivity is observed in the vicinity of strained pyramidal sector boundaries B_1. The growth band *GB* is the result of a relatively large and short temperature disturbance in the growth solution (about 1 °C in one hour). In the EDT photograph this band can be seen only at the inclusions *I* occurring in strained areas, such as dislocations *D* and the unions between growth sectors. B_2 boundaries between the prismatic and pyramidal sectors.

consist of a single pyramidal facet. The reason is simply that nucleation on such seeds has not yet proved entirely successful. Experiments in that direction led to the conclusion that in such a case the lattice parameters of the prospective crystal should differ as little as possible from those of the seed. Translated into the deuterium content x, this means that the difference in x should not exceed 0.002. This implies that the seed must have been grown from a solution having a deuterium content practically identical with that of the growth solution. Given this severe constraint and the need to achieve growth on both sides of the seed plate to increase the crystal production yield, seeding on (001) plates was finally retained as the most suitable process.

The initial growth step is therefore aimed at letting the crystal develop its natural habit, as shown in fig. 2. The cap is formed by the merging of the correct facets of pyramids that have nucleated on the (001) slice. In this process the occurrence of crystal defects, such as dislocations and mother-liquor inclusions, can be attributed to macroscopic or microscopic deformations of the seed plate, irregularities during the growth of the pyramids (elemental caps), and the merging of imperfectly coherent facets of two adjacent pyramids [16]. An example of the last two situations is shown in *fig. 13*. The (011) sector of

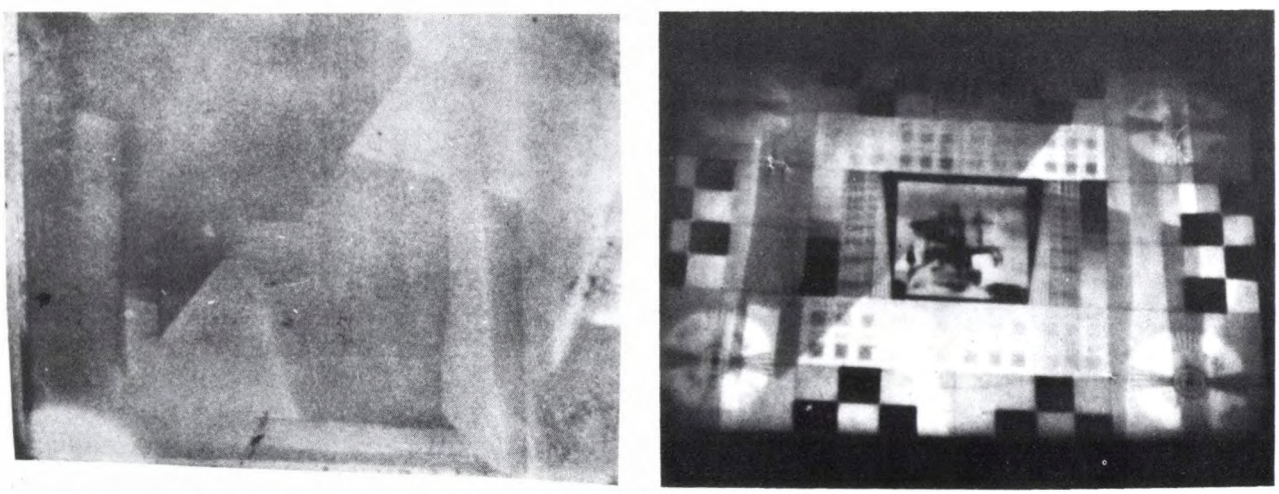


Fig. 12. EDT photograph (*left*) and TITUS image (*right*) obtained from two consecutive (001) slices of a DKDP crystal. These is a clear one-to-one relation between the locations of the heterogeneities in the electrical conductivity and the contrasts in the corresponding TITUS image.

pyramid *A* extents at the expense of the $(0\bar{1}1)$ sector of the same pyramid, and dislocation lines are generated at the union between the two sectors (arrow *1*). Dislocation lines also occur when two similar (011) sectors of the adjacent pyramids *A* and *B* merge (arrow *2*).

Non-ideal seeding conditions are responsible for the large strains across the pyramidal sectors, which are associated with large conductivity heterogeneities and photoelastic flaws in TITUS images [16]. These heterogeneities are usually found to decrease slowly along the direction of growth. Non-ideal seeding conditions are also responsible for elastic deformations at the boundaries between sectors that are otherwise coherent. These deformations, which produce dynamic contrasts in the X-ray topographs, may be associated with conductivity heterogeneities eventually stretching over a large distance in one sector. A third consequence of non-ideal seeding is the occurrence of the dislocation lines mentioned above [16,17,18]. These have an adverse effect on the crystal perfection because of their influence on the pyramidal growth rate and their interaction with strained sector boundaries, as we shall shortly see.

Although the formation of the cap is relatively fortuitous, the crystal damage caused during this stage can be kept surprisingly low by taking some simple precautionary measures; see *fig. 14*. When a relatively thick seed plate is used, deformation of the seed is avoided so that it forms a perfectly flat base for the growing cap. This gives the merging growth pyramids a high degree of coherence. By keeping the supersaturation in the growth solution very low at the start of the seeding process, the number of elemental caps remains limited. The use of thick seed plates combined with a limited number of elemental caps is found to give a dramatic reduction of these strains in the sectors and at the sector boundaries. At the same time the dislocation density is considerably reduced.

These findings have led to the growth procedure described above, with the additional requirement that during the early step conducted at very low supersaturation the elemental caps should not taper (see below). An increasing need for more perfect crystals for ever more exacting applications will presumably require the use of pyramidal seeds that are perfectly coherent with the crystal to be grown.

[16] C. Belouet, M. Monnier, E. Dunia and J. F. Pétroff, Mat. Res. Bull. **11**, 903, 1976.

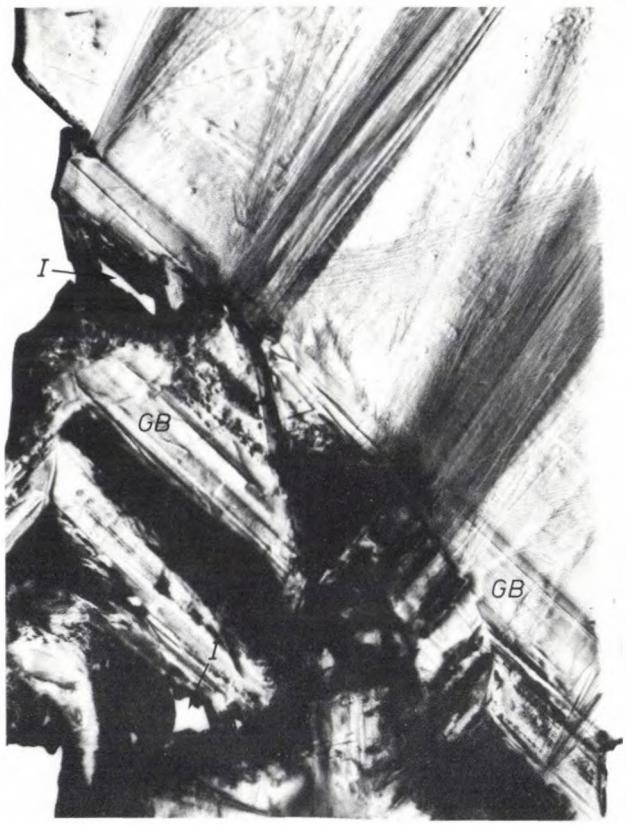
[17] Y. Epelboin, A. Zarka and H. Klapper, J. Crystal Growth **20**, 103, 1973.
V. G. Lutsau, Ju. M. Fishman and I. S. Res, Kristall und Technik **5**, 445, 1970.

[18] C. Belouet, E. Dunia and J. F. Pétroff, J. Crystal Growth **23**, 243, 1974.
C. Belouet, Rev. Phys. appl. **11**, 253, 1976.

Growth-induced defects

In most cases the defects that occur during growth can be related to the presence of impurities in the prismatic and pyramidal sectors. The incorporation of these impurities is found to depend closely on the stirring of the growth solution around the crystal and on temperature deviations in the solution.

Stirring plays an important part in the tapering of the prismatic facets along the [001] axis. The effect of tapering, which is always encountered in DKDP crystals, is attributed to poisoning of prismatic facet edges owing to the incorporation of impurities, usually hydrolysed iron and chromium complexes. As can be seen in *fig. 15*, this leads to a reduction of the rate of growth in the pyramidal sectors. This observation can



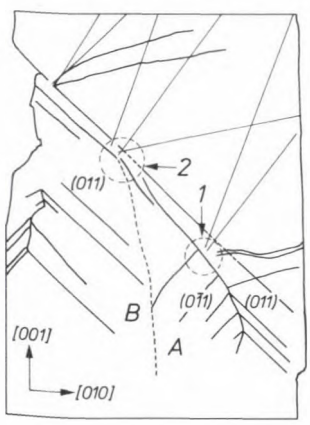


Fig. 13. Transverse X-ray topograph (*above*) of a (100) slice sawn from the grown tetragonal cap of a DKDP crystal, and schematic representation of the cap formation (*right*). Dislocation lines are generated if the (011) and the $(0\bar{1}1)$ sectors of the same pyramid have different lattice constants (arrow *1*), or if the two merging (011) sectors of the adjacent pyramids *A* and *B* are at an angle to each other (arrow *2*). The topograph also shows the inclusions *I* and the growth bands *GB* mentioned ealier.

be explained on the assumption that the rate of pyramidal growth is determined by nucleation. In this hypothesis the strain associated with tapering reduces the rate of generation of nuclei at the intersection between the prismatic and pyramidal sectors [14]. Because of the tapering effect the likelihood of undesirable overlapping of adjacent pyramidal sectors increases, and crystal growth becomes more sensitive to temperature deviations in the growth solution. The degree of tapering depends on the supersaturation and the *p*H of the growth solution, and also on the concentration of certain elements, particularly iron in the present experimental conditions. *Fig. 16* gives an illustrative example of the influence of the iron concentration in the mother liquor. A deliberate large increase in the iron content causes a sharp increase in the angle of taper. Experiments with the vessel shown in fig. 5 showed that the segregation of iron in the prismatic sectors, which determines the degree of taper, depends closely on the rate of rotation ω of the crystal about the [001] axis and the diagonal length d of the (001) slice. From the results of the measurements a relation is deduced between the minimum iron content — or the minimum taper angle — and an optimum value of the product ωd^2. With a suitable stirring rate, it was found possible to obtain non-tapered crystals (*fig. 17*) with no deformations at sector boundaries, provided that the iron content in the growth solution was kept below 1 ppm. Applied to large growth chambers [6] these results point to the necessity of vigorously stirring the growth solution with the rotation-reversal system mentioned earlier.

Temperature deviations in the growth solution cause changes in the prismatic and the pyramidal growth rates, which in turn affect the incorporation of impurities in these sectors. This process gives rise to the formation of growth bands, which can be observed in the X-ray topographs from their slightly different lattice constants, and in the EDT photographs from their different electrical conductivity (fig. 11). The increased incorporation of iron, for example, at the sites of a growth band in a pyramidal sector can be demonstrated by comparing autoradiographs of crystals grown from a ^{59}Fe-doped solution with the corresponding X-ray topographs; see *fig. 18*. The changes in iron content, correlated with the growth parameters, are found to be much greater for the prismatic sectors than for the pyramidal ones. A comparison of the densitograms of the autoradiographs and the X-ray topographs shows a one-to-one correspondence between the location of the iron bands in the prismatic sectors and the contrasts (growth bands) in the topographs; see *fig. 19*. The effectiveness with which temperature deviations cause growth

Fig. 14. X-ray topograph of a (001) slice of a DKDP crystal sawn off close to the seed crystal before the final growth habit has been attained. Observe the continuity of the '*Pendellösung*' fringes at the boundary between the pyramidal sectors. Fringe deformations are due to thickness variations.

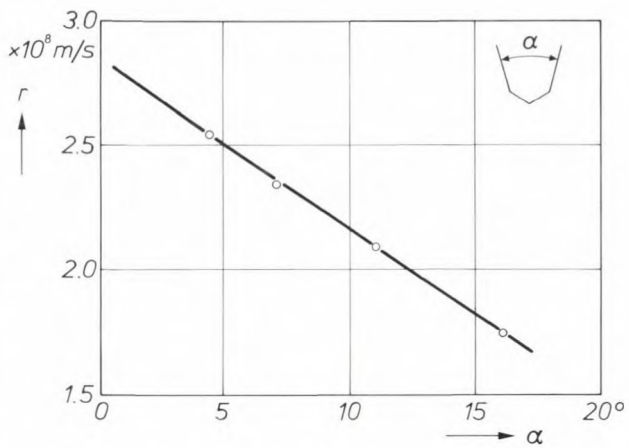


Fig. 15. Ratio r of the growth rate of the pyramidal facets of a DKDP crystal to the supersaturation in the growth solution at about 36 °C, plotted as a function of the angle of taper α. With supersaturation and temperature constant, the pyramidal growth rate decreases linearly with α.

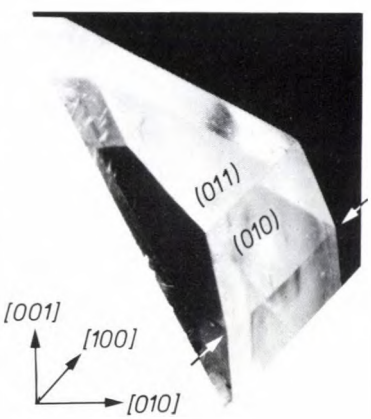


Fig. 16. Photograph of part of a tapered DKDP crystal. The abrupt change in the angle of taper (indicated by an arrow) is the result of deliberately increasing the iron content in the growth solution from 1 to 10 ppm, other conditions for growth being maintained.

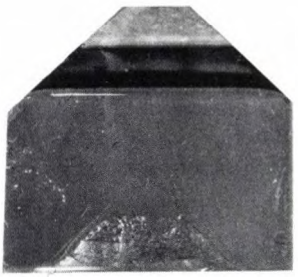


Fig. 17. Photograph of a non-tapered DKDP crystal. The initial stage in the forming of the tetragonal cap can clearly be seen. The edges at the corners of the facets are somewhat rounded because of the use of a protective film of oil.

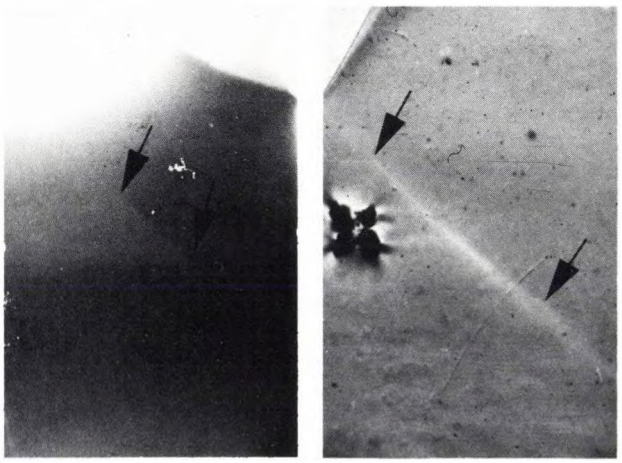


Fig. 18. Autoradiograph (*left*) and X-ray topograph (*right*) of a (001) slice of a KH_2PO_4 (KDP) crystal grown from a solution doped with ^{59}Fe. The growth band in a pyramidal sector is indicated in both cases with arrows.

The growth band *GB* shown in fig. 11 was the result of a rapid temperature disturbance of about 1 °C. The band appears as a quasi-planar fault that can be detected in the EDT photographs only at locations where the crystal was initially strained. Increased changes in the concentration of impurities and in the electrical conductivity in response to temperature fluctuations are frequently observed in strained areas, more particularly near the tip of the cap and at the boundaries between adjacent interacting sectors. This effect gives rise to the non-continuous growth bands found in DKDP crystals, and to the occurrence of imperfections such as growth-solution inclusions.

The maximum permissible temperature fluctuation requires very accurate temperature control, very good insulation of the growth chamber and a large volume of growth solution. Routine growth of DKDP crystals is now possible with maximum temperature fluctuations of about 0.005 °C. Any growth bands that might then still be present give such a small change in the electrical conductivity that they have no adverse effect on picture reproduction with the TITUS tube.

Interaction of crystal defects

A troublesome effect is found to be associated with the 'refraction' of a dislocation line at the boundary between two adjacent pyramidal sectors. When a dislocation line (of the type marked with arrow *2* in fig. 13 in all cases investigated) of a (101) sector, inclined in the direction of the [001] axis in a (010)

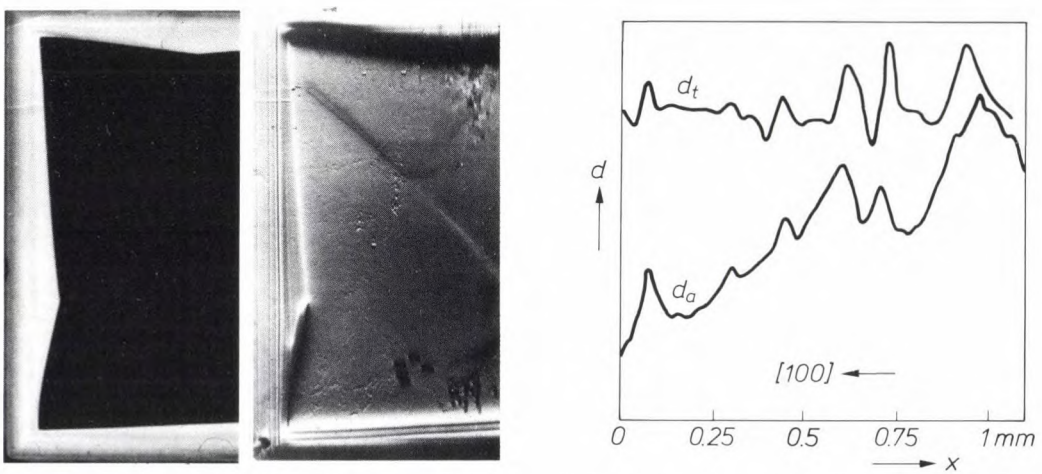


Fig. 19. Autoradiograph (*left*) and X-ray topograph (*right*) of a (001) slice of a KDP crystal grown from a solution doped with ^{59}Fe. The autoradiograph shows that the iron content in the prismatic sectors is much higher than in the pyramidal sectors (a white area indicates the presence of a relatively large amount of iron). The densitograms d_a and d_t were obtained by scanning corresponding locations x in a prismatic sector in the autoradiograph and the topograph with a microdensitometer. Observe the one-to-one correspondence between the location of the iron bands and the contrasts in the topograph.

bands is determined by the degree of taper, the impurity content of the growth solution, the strains in the growth sectors and the deuterium content. Non-tapered crystals grown from very pure solutions (*Table II*) have hardly any growth bands if the temperature deviations are less than 0.01 °C.

plane, happens to be cut off by the boundary of a slowly growing (011) sector, then the line is 'refracted' on entering this sector, as illustrated in *fig. 20.* At the same time three quasi-planar defects *A, B* and *C* are generated in the (011) sector. The defects *A* and *B* intersect at the dislocation line; *B* extends over the

whole sector, giving rise to distinctly different contrasts on either side of A in the X-ray topographs. Analysis of these topographs suggests that these defects are not purely planar and that no discrete deformation vectors can be assigned to any of the interacting defects. The pyramid-like region delineated by the boundary between the (010) and (011) sectors and the defects A, B and C extend in the growth direction as the dislocation line propagates.

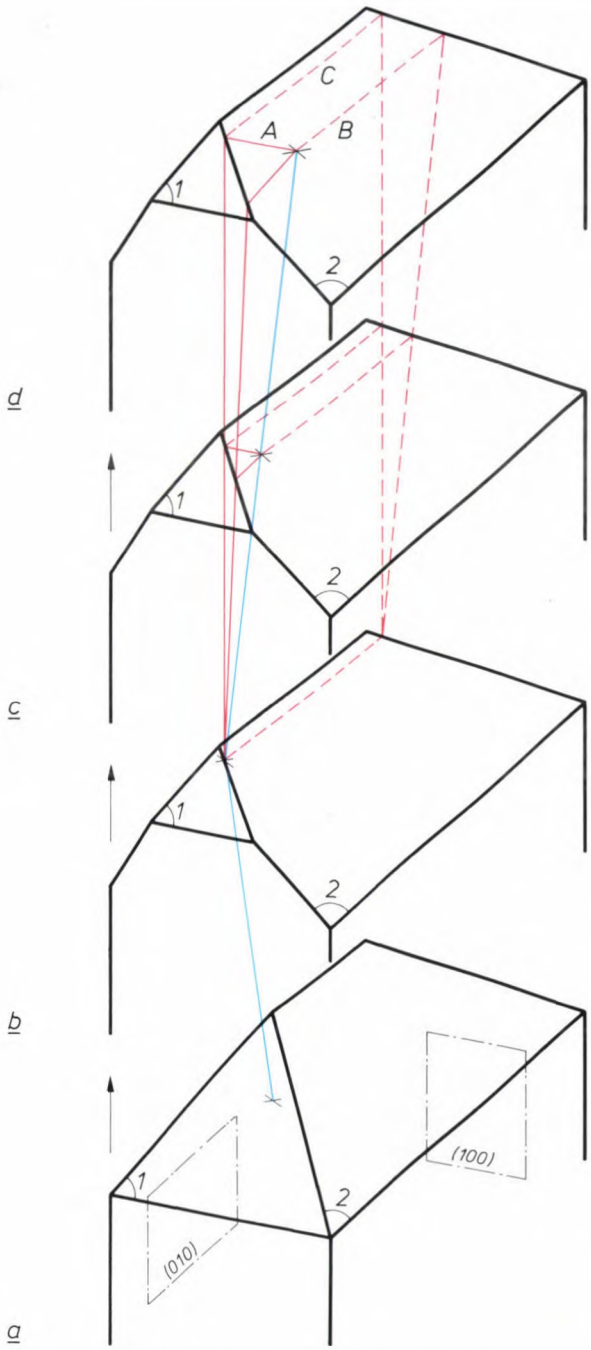


Fig. 20. Diagram illustrating a mechanism that gives rise to a new bulk defect in the (011) sector because of the refraction of a dislocation line (*blue*) at the boundary between a (101) sector (*1*) and a (011) sector (*2*). Sector *1* is assumed to have a more rapid growth up to the step (*b*) where the dislocation line enters sector *2*. From (*b*) to (*d*) the two sectors have the same growth rate. The bulk defect thus induced in sector *2* extends throughout the sector delineated by the nearly planar boundaries *A*, *B* and *C* (*solid and dashed red lines*) and grows as the dislocation line propagates. In the region delineated by red lines the electrical conductivity is higher or lower than in the rest of sector *2*, owing to the generated defect. The crosses indicate the emergence of the dislocation line on the facets.

Table II. Concentrations of a number of elements in a mother liquor permitting the growth of high-quality DKDP crystals.

Element	Concentration (ppm)	Element	Concentration (ppm)
Barium	0.08	Iron	<1.0
Calcium	0.7	Rubidium	5
Cerium	0.011	Sulphur	<1.0
Cobalt	0.0066	Antimony	0.17
Chromium	0.0077	Strontium	0.10
Caesium	0.0021	Thorium	0.0059
Copper	<1.0	Zinc	0.59

The areas on both sides of A may have quite a different conductivity, as can be seen for example in fig. 10. Comparison of the EDT photographs with the X-ray topographs reveals that the magnitude of the conductivity change is closely related to the magnitude of the strain at the boundary between the interacting sectors. This points to an interaction of defects when a dislocation line is 'refracted' by a strained sector boundary, giving rise to additional heterogeneities in the electrical conductivity. This effect is found only when adjacent pyramidal growth sectors can overlap. To obtain homogeneous crystals it is therefore essential to eliminate or at least minimize the various causes of overlapping. At the same time the occurrence of excessive strains at the union between pyramidal sectors should be prevented.

Finally, *Table III* presents a summary of the most important precautionary measures that should be taken for the reproducible growth of DKDP crystals that will give good performance in the TITUS tube. The high quality of such crystals can be seen from X-

Fig. 21. X-ray section topograph of a 1 mm thick DKDP slice. The regular fringes demonstrate the perfection of the crystal. This topograph was made by E. Dunia.

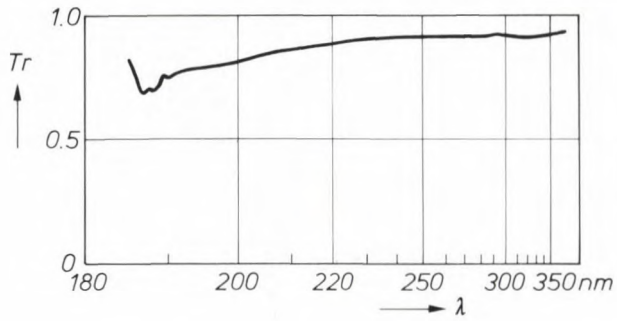


Fig. 22. Ultraviolet transmission *Tr* as a function of the wavelength λ of a 20 mm thick DKDP crystal of high quality. Unlike most of the DKDP crystals described so far, the crystal shows no absorption band at about 300 nm.

Table III. Precautionary measures for growing high-quality DKDP crystals.

Parameter	Precautions
Form of seed crystal	(001) plate
Thickness of seed	More than 5 mm
Deuterium content in the seed	In equilibrium with deuterium content in the growth solution
Cap formation	Very low supersaturation in the growth solution
Stirring	Optimum speed of growth-solution rotation around the crystal
Impurities	Minimum content, especially of [Fe] and [Cr] $\ll$ 1 ppm
Temperature in growth solution	Deviations less than 0.01 °C

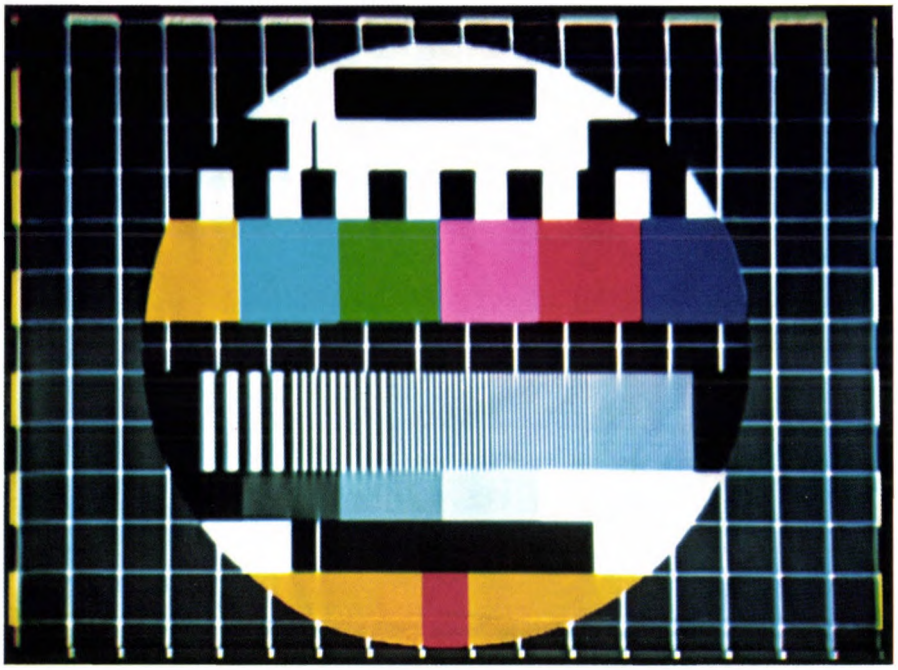


Fig. 23. Photograph of a test pattern projected on to a screen of 8 m² by means of a TITUS tube using a high-quality DKDP crystal. The maximum screen illumination is 350 lux. The resolution pattern shows 104, 208, 312, 416 and 520 lines over the effective picture width. The maximum resolution is 750 lines.

ray topographs (see for example *fig. 21*) and from the absence of an absorption band at about 300 nm (*fig. 22*). When these crystals are used there are no conductivity heterogeneities in the TITUS pictures, even after decay periods of longer than an hour. *Fig. 23* shows a test pattern projected on a screen from a TITUS tube with a DKDP crystal target of high quality.

Summary. Single crystals of deuterated potassium dihydrogen phosphate (DKDP) are used as a solid-state light modulator in the TITUS tube for the large-screen projection of television pictures. The crystals are tetragonal and form a four-sided prism with a pyramidal cap on both sides. They are grown on (001) seed plates from supersaturated aqueous solutions, preferably by the temperature-decrease method. X-ray topographs indicate the presence of various defects, such as dislocations, strains, growth bands and inclusions of mother liquor. These defects have a detrimental effect on the displayed pictures because they affect the electro-optical properties of the crystal. By combining these observations with the results of chemical analyses, neutron activation, autoradiography and electrode decoration it is shown that the defects are related to the properties of the seed plate and mother liquor and to temperature deviations during growth. The occurrence of defects can be largely suppressed by the use of a relatively thick seed plate, adjustment of the deuterium concentrations in seed and liquor, a very low supersaturation in the capping phase of growth, stirring the growth solution around the crystal at an optimum speed, a very low impurity content, and finally very good temperature control.

Scientific publications

These publications are contributed by staff of laboratories and plants which form part of or cooperate with enterprises of the Philips group of companies, particularly by staff of the following research laboratories:

Philips Research Laboratories, Eindhoven, The Netherlands	*E*
Philips Research Laboratories, Redhill, Surrey, England	*R*
Laboratoires d'Electronique et de Physique Appliquée, 3 avenue Descartes, 94450 Limeil-Brévannes, France	*L*
Philips GmbH Forschungslaboratorium Aachen, Weißhausstraße, 51 Aachen, Germany	*A*
Philips GmbH Forschungslaboratorium Hamburg, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany	*H*
Philips Research Laboratory Brussels, 2 avenue Van Becelaere, 1170 Brussels (Boitsfort), Belgium	*B*
Philips Laboratories, N.A.P.C., 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A.	*N*

E. Arnold: Comment on the frequency dependence of electron conductivity in the silicon inversion layer in the metallic and localized regimes.
Phys. Rev. B **17**, 4111-4113, 1978 (No. 10). *N*

M. O. Bargouth, G. Will (both with Universität Bonn) **& K. H. J. Buschow:** The magnetic structure of $ErFe_4Al_8$ and $TbFe_4Al_8$.
J. Magn. magn. Mat. **6**, 129-130, 1977. *E*

C. I. M. Beenakker & P. W. J. M. Boumans: Additional experience with the cylindrical TM_{010} cavity for generating an MIP in helium and argon at atmospheric pressure.
Spectrochim. Acta **33B**, 53-54, 1978 (No. 1/2). *E*

C. Belouet, J. J. Brissot, R. Martres & Ngo Tich Phuoc: Continuous polycrystalline silicon layers on carbon substrates.
Photovoltaic Solar Energy Conf., Luxembourg 1977, pp. 164-175; 1978. *L*

Th. Berben & C. Büthker (Philips Elcoma Division, Eindhoven): Isostatic hot-pressing and its influence on microstructure.
Science of Ceramics **9**, 176-182, 1977.

P. Blood, L. C. Feldman*, G. L. Miller* & J. P. Remeika* (* Bell Laboratories, Murray Hill, N.J.): The rotating sample technique for measurement of random backscattering yields from crystals and its application to β-alumina.
Nucl. Instr. Meth. **149**, 225-228, 1978 (No. 1-3). *R*

A. J. van Bommel, J. E. Crombeen & T. G. J. van Oirschot: LEED, AES and photoemission measurements of epitaxially grown GaAs(001), (111)A and $(\bar{1}\bar{1}\bar{1})$B surfaces and their behaviour upon Cs adsorption.
Surface Sci. **72**, 95-108, 1978 (No. 1). *E*

P. Bordewijk (University of Leiden) **& W. H. de Jeu:** Calculation of dipole correlation factors in liquid crystals with use of a semiempirical expression for the internal field.
J. chem. Phys. **68**, 116-118, 1978 (No. 1). *E*

H. Bosma: How to master complexity: decomplexing. Real time programming 1977, Proc. IFAC/IFIP Workshop, Eindhoven, ed. C. H. Smedema, pp. xi-xiv; Pergamon Press, Oxford 1978. *E*

D. J. Breed, W. T. Stacy, A. B. Voermans, H. Logmans & A. M. J. van der Heijden: New bubble materials with high peak velocity.
IEEE Trans. **MAG-13**, 1087-1091, 1977 (No. 5). *E*

A. Broese van Groenou: Experiments on grinding of ceramic ferrites.
Science of Ceramics **9**, 486-493, 1977. *E*

H. H. Brongersma, L. C. M. Beirens & G. C. J. van der Ligt: Applications of low-energy ion scattering.
Material characterization using ion beams, ed. J. P. Thomas & A. Cachard, pp. 65-79; Plenum Press, New York 1978. *E*

H. H. Brongersma, M. J. Sparnaay & T. M. Buck (Bell Laboratories, Murray Hill, N.J.): Surface segregation in Cu-Ni and Cu-Pt alloys; a comparison of low-energy ion-scattering results with theory.
Surface Sci. **71**, 657-678, 1978 (No. 3). *E*

S. D. Brotherton & J. Bicknell: The electron capture cross section and energy level of the gold acceptor center in silicon.
J. appl. Phys. **49**, 667-671, 1978 (No. 2). *R*

E. Bruninx: Radiochemistry at the Philips Research Laboratories, Eindhoven, The Netherlands.
J. radioanal. Chem. **40**, 233-235, 1977 (No. 1/2). *E*

R. Bruno: A correction procedure for separating direct and diffuse insolation on a horizontal surface.
Solar Energy **20**, 97-100, 1978 (No. 1). *A*

F. M. A. Carpay: Growth of composites by eutectoid or other solid-state decompositions.
Int. Metals Revs **23**, 1-18, 1978 (No. 1). *E*

F. M. A. Carpay & S. K. Kurtz: A multiple-lognormal model of normal grain growth.
Science of Ceramics **9**, 92-102B, 1977. *E*

M. Cathelin & G. Durand: Circuits intégrés logiques à transistors à effet de champ sur GaAs en technologie plane.
Onde électr. **58**, 218-221, 1978 (No. 3). *L*

C. K. Davis & R. F. Mitchell: Studies of small trunking systems for mobile radio.
Conf. on Communications equipment and systems, Birmingham 1978 (IEE Conf. Publn No. 162), pp. 143-146. *R*

P. Delsarte: Partial-optimal piecewise decoding of linear codes.
IEEE Trans. **IT-24**, 70-75, 1978 (No. 1). *B*

P. Delsarte, Y. V. Genin & Y. G. Kamp: Orthogonal polynomial matrices on the unit circle.
IEEE Trans. **CAS-25**, 149-160, 1978 (No. 3). *B*

P. A. Devijver: Statistical methods of pattern recognition.
Data structure analysis and applications, ed. J. Kittler, pp. 46-79; ENST-C-78002, Paris 1978. *B*

G. Diefenbach & H. Schemmann: Rotierender Antrieb für ein Gerät mit oszillierender Arbeitsbewegung.
Elektrotechn. Z. etz.b **30**, 56-60, 1978 (No. 2). *A*

R. J. Dolphin & F. W. Willmott: Separation of chlorinated dibenzo-*p*-dioxins from chlorinated congeners.
J. Chromatography **149**, 161-168, 1978. *R*

E. Dormann, M. Huck (both with T. H. Darmstadt) **& K. H. J. Buschow:** NMR in Laves phase alloy system $GdPt_x$ — analysis of nonstoichiometry in intermetallic compounds.
Z. Physik B **27**, 141-151, 1977 (No. 2). *E*

H. Dötsch, P. Röschmann & W. Schilz: Ferrimagnetic resonance spectra of magnetic bubble films at low microwave frequencies.
Appl. Phys. **15**, 167-173, 1978 (No. 2). *H*

H. Durand: Perspectives et limitations économiques de l'emploi de la conversion photovoltaïque.
Session de formation 'Applications terrestres des cellules solaires photovoltaïques', Ecole des Mines de Madrid 1978, 41 pp. *L*

E. Fabre & Y. Baudet: Polycrystalline silicon solar cells.
Photovoltaic Solar Energy Conf., Luxembourg 1977, pp. 178-186; 1978. *L*

E. Fabre, A. Brière & J. P. André: Metal alkyl grown GaAs solar cells.
Photovoltaic Solar Energy Conf., Luxembourg 1977, pp. 388-394; 1978. *L*

E. Fabre, L. De Smet* & R. Mertens* (* Katholieke Universiteit, Leuven): High intensity silicon solar cells.
Photovoltaic Solar Energy Conf., Luxembourg 1977, pp. 249-258; 1978. *L*

A. Flett: Adjacent channel performance of FM, AM and SSB at VHF.
Conf. on Communications equipment and systems, Birmingham 1978 (IEE Conf. Publn No. 162), pp. 137-139. *R*

J. C. Gijsbers & J. G. Kloosterboer: Kinetic determination of borate at the parts per million level.
Anal. Chem. **50**, 455-457, 1978 (No. 3). *E*

J. J. Goedbloed & E. T. J. M. Smeets: Very low noise silicon planar avalanche photodiodes.
Electronics Letters **14**, 67-69, 1978 (No. 3). *E*

J. L. Harthoorn: Ultrasonic metal welding.
Thesis, Eindhoven 1978. *E*

J. P. Hazan: Intensity profile distortion due to resolution limitation in fibre index profile determination by near field.
Electronics Letters **14**, 158-160, 1978 (No. 5). *L*

H. Heitmann, C. Fritzsche (Inst. f. Angew. Festkörperphysik, Freiburg), **P. Hansen, J.-P. Krumme, R. Spohr** (Ges. f. Schwerionenforschung, Darmstadt) **& K. Witter:** Influence of irradiation with high energetic ions on storage properties of magneto-optic $(Gd,Bi)_3(Fe,Ga)_5O_{12}$ epitaxial films.
J. Magn. magn. Mat. **7**, 40-43, 1978 (No. 1-4). *H, E*

H. Heyns: Resistive-gate CTD area image sensor.
1978 IEEE Int. Solid-State Circuits Conf. Dig. tech. Papers, pp. 32-33 + 262. *E*

A. Huijser, J. van Laar & T. L. van Rooy: Angle-resolved photoemission from GaAs(110) surface states.
Physics Letters **65A**, 337-339, 1978 (No. 4). *E*

L. Jacomme: Wavelength dependent propagation in optical multimode fibers.
Optics Comm. **24**, 213-216, 1978 (No. 2). *L*

W. H. de Jeu (I, II), **W. A. P. Claassen** (I) **& P. Bordewijk** (University of Leiden) (II): Physical studies of nematic azoxybenzenes: I. Magnetic susceptibilities and the order parameter, II. Refractive indices and the internal field.
J. chem. Phys. **68**, 102-108, 109-115, 1978 (No. 1). *E*

F. M. Klaassen: Design and performance of micron-size devices.
Solid-State Electronics **21**, 565-571, 1978 (No. 3). *E*

J. Köhler & B. Schiek: Broadband microwave frequency doublers.
Radio and electronic Engr. **48**, 29-32, 1978 (No. 1/2). *H*

R. Koppe: Automatische Abbildung eines planaren Graphen in die Ebene mit beliebig vorgebbaren Örtern der Knotenbilder.
Computing **20**, 61-73, 1978 (No. 1). *H*

E. Krätzig & R. Orlowski: $LiTaO_3$ as holographic storage material.
Appl. Phys. **15**, 133-139, 1978 (No. 2). *H*

H. K. Kuiken: Heat or mass transfer from an open cavity.
J. Engng. Math. **12**, 129-155, 1978 (No. 2). *E*

J. Lohstroh & A. Slob: Punch-through cell for dense bipolar ROMs.
1978 IEEE Int. Solid-State Circuits Conf. Dig. tech. Papers, pp. 20-21. *E*

G. M. Loiacono, M. Delfino & W. A. Smith: Specific heats of NdP_5O_{14} and PrP_5O_{14} near their ferroelastic phase transitions.
Appl. Phys. Letters **32**, 595-596, 1978 (No. 10). *N*

K. M. Lüdeke, B. Schiek & J. Köhler: Radiation balance microwave thermograph for industrial and medical applications.
Electronics Letters **14**, 194-196, 1978 (No. 6). *H*

J. Magarshack: FETs: improving performance and reliability.
Microwave Syst. News **8**, No. 2, pp. 38 + 40, Feb. 1978. *L*

A. B. D. van der Meer (Philips Elcoma Division, Eindhoven): Strength measurements on strontium hexaferrites.
Science of Ceramics **9**, 535-542, 1977.

J. Michel & F. Lasnier: Electrical characterization of silicon diffused layers.
Photovoltaic Solar Energy Conf., Luxembourg 1977, pp. 125-134; 1978. *L*

A. R. Miedema, F. R. de Boer (Universiteit van Amsterdam) **& R. Boom** (Hoogovens, IJmuiden): Model predictions for the enthalpy of formation of transition metal alloys.
Calphad **1**, 341-359, 1977 (No. 4). *E*

A. R. Miedema & R. Boom (Hoogovens, IJmuiden): Surface tension and electron density of pure liquid metals.
Z. Metallkunde **69**, 183-190, 1978 (No. 3). *E*

G. J. Naaijer: Putting to use solar energy with silicon photovoltaic cells.
Photovoltaic Solar Energy Conf., Luxembourg 1977, pp. 809-816; 1978. *L*

J. Paredaens: On the expressive power of the relational algebra.
Inform. Process. Letters **7**, 107-111, 1978 (No. 2). *B*

J. G. J. Peelen, B. V. Rejda*, J. P. W. Vermeiden* & K. de Groot* (* Free University of Amsterdam): Sintered tricalciumphosphate as bioceramic.
Science of Ceramics **9**, 226-236, 1977. *E*

H. I. Ralph: The degeneracy factor of the gold acceptor level in silicon.
J. appl. Phys. **49**, 672-675, 1978 (No. 2). *R*

H. Rau: Defect energetics and range of homogeneity of α-MnS.
J. Phys. Chem. Solids **39**, 339-343, 1978 (No. 4). *A*

T. E. Rozzi, J. H. C. van Heuven & G. H. in 't Veld: A new d.h. laser configuration with passive transverse field confinement.
Electronics Letters **14**, 87-88, 1978 (No. 4). *E*

S. Strijbos & A. C. Knaapen: Mechanical properties of a ferrite powder and its granulate.
Science of Ceramics **9**, 477-485, 1977. *E*

M. D. Sturge (Bell Laboratories, Murray Hill, N.J.), **A. T. Vink & F. P. J. Kuijpers:** A new interimpurity recombination in GaP; revised values for acceptor binding energies.
Appl. Phys. Letters **32**, 49-51, 1978 (No. 1). *E*

Y. Tamminga & W. J. M. J. Josquin: The regrowth of amorphous layers created by high-dose antimony implantation in ⟨100⟩ silicon.
Appl. Phys. Letters **32**, 13-15, 1978 (No. 1). *E*

H. J. L. Trap: L'emploi des verres au germano-silicate dans la passivation des dispositifs à semi-conducteurs.
Verres Réfract. **32**, 17-23, 1978 (No. 1). *E*

M. J. Underhill: Universal frequency synthesiser I.C. system.
Conf. on Communications equipment and systems, Birmingham 1978 (IEE Conf. Publn No. 162), pp. 185-188. *R*

M. J. Underhill: Transient and frequency responses of systems with time delays.
Electronics Letters **14**, 284-286, 1978 (No. 9). *R*

J. D. B. Veldkamp: Crack formation during scratching.
Science of Ceramics **9**, 438-446, 1977. *E*

J. D. B. Veldkamp, N. Hattu & V. A. C. Snijders (College of Advanced Technology, Utrecht): Crack formation during scratching of brittle materials.
Fracture mechanics of ceramics, Vol. 3, ed. R. C. Bradt, D. P. H. Hasselman & F. F. Lange, pp. 273-301; Plenum Press, New York 1978. *E*

W. F. van der Weg (Philips Res. Labs, Amsterdam Division): Ion beam induced light emission: mechanisms and analytical applications.
Material characterization using ion beams, ed. J. P. Thomas & A. Cachard, pp. 81-99; Plenum Press, New York 1978.

H. W. Werner: Round table on comparison of secondary ion mass spectrometry with other analytical methods.
Adv. in Mass Spectrometry **7A**, 804-806, 1978. *E*

H. W. Werner & A. E. Morgan: Some considerations regarding the applicability of the Saha-Eggert equation to quantitative SIMS analysis.
Adv. in Mass Spectrometry **7A**, 764-769, 1978. *E*

P. D. White, R. Stevens & R. F. Mitchell: Surface acoustic wave resonators in communications.
Conf. on Communications equipment and systems, Birmingham 1978 (IEE Conf. Publn No. 162), pp. 270-273. *R*

G. Will, M. O. Bargouth (both with Universität Bonn) **& K. H. J. Buschow:** The antiferromagnetic structure of $TbMg_3$.
J. Magn. magn. Mat. **6**, 131-133, 1977. *E*

G. Will, V. Lehmann (both with Universität Bonn) **& K. H. J. Buschow:** Neutron diffraction investigation of ferromagnetic TbB_2.
J. Magn. magn. Mat. **6**, 22-23, 1977. *E*

Supplement to Philips Technical Review
Beilage der Philips Technischen Rundschau
Bijlage van Philips technisch Tijdschrift

PHILIPS

August 1980

Recent United States Patents

Abstracts from patents that describe inventions from the following research laboratories that form part of or cooperate with the Philips group of companies:

Philips Research Laboratories, Eindhoven, The Netherlands	*E*
Philips Research Laboratories, Redhill, Surrey, England	*R*
Laboratoires d'Electronique et de Physique Appliquée, 3 avenue Descartes, 94450 Limeil-Brévannes, France	*L*
Philips GmbH Forschungslaboratorium Aachen, Weißhausstraße, 51 Aachen, Germany	*A*
Philips GmbH Forschungslaboratorium Hamburg, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany	*H*
Philips Research Laboratory Brussels, 2 avenue Van Becelaere, 1170 Brussels (Boitsfort), Belgium	*B*
Philips Laboratories, N.A.P.C., 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A.	*N*

4 166 969

Target and target assembly for a camera tube and method of manufacturing same

A. M. E. Hoeberechts *E*

A target and target assembly for a camera tube in which a semiconductor plate is provided on an annular support and consists of a semiconductor monocrystalline edge portion which comprises an integrated circuit for handling the electrical signals originating from the target, the central part consisting of a radiation-sensitive layer having one or more radiation-permeable electrodes. The electrodes are connected to inputs of the integrated circuit of which the leads necessary for supply and control voltages are led through. A window is provided on the electrodes and overlaps the inner edge of the support, the window, the edge portion and the support adjoining each other in a vacuum-tight manner.

4 167 035

Voltage multipliers

G. A. Willard *R*

A voltage multiplier of the Cockroft-Walton type in which the a.c.- and d.c.-operated capacitors are arranged in respective separate capacitor winding rolls, the two rolls being arranged side by side on a common axis and the diodes being arranged between the two rolls.

4 167 039

Device for measuring the distribution of the absorption of the emission of radiation in a layer of a body

G. Kowalski *H*

In devices for reconstructing the adsorption distribution in a layer of a body on the basis of the measuring values obtained by means of a CT scanner, the measuring values (or values derived therefrom) are 'spread' along strips which occupy the same position relative to the image plane as the strips along which the measuring values were obtained. The invention assigns the measuring values to individual image elements, the calculation of weighting factors being eliminated by subdividing each image element into smaller image points (9 to 16 times smaller).

4 167 758

Method and device for transmitting facsimile picture signals

U. Rothgordt *H*

With the facsimile transmission of pictures the run-lengths (plurality of successive picture elements of the same luminance value) are often transmitted in coded form to save transmission time and transmission bandwidth respectively. Herein the shortest run-length is often allocated to the shortest code word. In accordance with the invention every run-length having the value 1 is either extended to the value 2 or suppressed and the shortest code word is then allocated in a corresponding manner to the run-length 2. As a result the total length of the transmitted code words for a document is considerably reduced. Extension is done such that one picture element is extended to the left to the detriment of the preceding run-length having the opposite luminance value provided this preceding run-length still contains thereafter at least two picture elements and that in the other case one picture element is extended to the right to the detriment of the next run-length provided the latter retains at least the value 2 and that otherwise the run-length having the value 1 is suppressed. It is true that the resolution is slightly reduced, however, only for the very thin lines which is fully acceptable and in many cases hardly noticeable.

4 168 213

Field emission device and method of forming same

A. M. E. Hoeberechts *E*

A field emission device and method of forming same, comprising a substrate on which at least one conical electrode is provided, which substrate, with the exception of the proximity of the tip of the electrode, is covered with a layer of a dielectric material on which a conductive layer is present at least locally, in which in order to form an integrated accelerating electrode the conductive layer extends in the direction of the punctiform tip of the electrode to beyond the dielectric layer and shows an aperture above the tip so that the conductive layer forms a cap-shaped accelerating electrode surrounding the conical electrode.

4 168 444

Imaging devices

J. G. van Santen *E*

An imaging device includes a semiconductor body and an array of photosensitive elements arranged in rows and columns at a first portion of the body. A plurality of charge-transfer lines interdigitated with the columns of the array are provided for transferring storage charge carriers from the array to a charge-transfer shift register at a second portion of the semi-conductor body. Each of the transfer lines includes an elongate resistor electrode which has connections for applying a potential difference along the resistive electrode to produce a drift field in the underlying body portion in the direction of the shift register for transporting mobile charge carriers toward the shift register.

4 168 632

Variable angle swashplate drive

H. Fokker E

A swashplate drive in which the plate is tiltable between a minimum angle $\check{\alpha}$ and a maximum angle $\hat{\alpha}$ and one or more drive rods each have a drive rod head with two spherical surfaces which are contacted by two sliding bodies, the other, flat side of the bodies engaging two running surfaces on the plate. The running surfaces are spaced apart a selected distance which minimizes variation in play between the plate, sliding bodies and drive rod head.

---

4 169 189

Magnetic structure

W. T. Stacy E
A. B. Voermans
H. Logmans

A magnetic structure for the propagation of magnetic bubbles at elevated velocity. A magnetic bubble layer is grown on a [110] face of a substrate, the lattice misfit being between -6×10^{-3} and -2×10^{-3}, and the magnetic layer having a composition on the basis of europium-iron garnet and a damping parameter $\leq 3\times10^{-7}$ Oe^2 sec/rad.

---

4 169 986

Television camera tube having resilient gauze supporting structure

J. Stelwagen E
A. J. T. van Sambeeck

If in a television camera tube the gauze-shaped electrode which is tensioned before the signal plate is assembled in a damped mass-spring system, in which the ratio of the mass of said system to the mass of the gauze should be larger than 30, while the fundamental frequency of the gauze and the natural frequency of the mass system differ less than 10% from each other and the damping of the mass spring system is so large that the time necessary for a decrease in amplitude of 20 db of the vibration of the mass of the mass spring system is smaller than 0.5 second, a tube is obtained having a low microphony.

---

4 171 029

Vehicle propulsion system with inertial storage

J. R. A. Beale R

A vehicle propulsion and regenerative braking arrangement providing multi-mode operation with an engine and/or energy storage flywheel coupled to an automatic transmission. An electronic control has driver-selectable control programs offering maximum economy, or enhanced accelerating power, or extra power. A novel clutch protects against flywheel over-speeding and full engine braking is provided when the flywheel is overcharged.

---

4 171 225

Electroless copper plating solutions

A. Molenaar E
H. M. van den Bogaert
J. Boven

Electroless copper plating bath containing a copper salt, a complexing agent, alkali and by way of reducing agent a complex of formaldehyde with an aminocarboxylic acid, -sulphonic acid or -phosphonic acid. Consequently, decomposition of the bath is greatly reduced and furthermore copper of a superior quality is obtained.

4 171 408

Device operating on the displacement of magnetic domain walls

F. H. de Leeuw E

In a device operating on the displacement of magnetic domain walls and comprising a layer of magnetizable material having an easy axis of magnetization normal to the plane of the layer, in which layer magnetic domain walls exist, the presence in the walls of an internal magnetic field directed parallel to the plane of the layer ensures that the walls can be displaced with higher velocity. Said internal field is created by providing on the first layer a second layer of magnetizable material having an easy axis of magnetization normal to the plane of the layer, the saturation magnetization of the second layer being higher than that of the first layer.

---

4 171 524

Display device having a matrix of gas discharge display elements

J. H. L. Lorteije E
G. Warrink
G. H. F. de Vries
J. A. M. Hulshof

In a matrix of display elements arranged in rows and columns having row scanning and column selection, all display elements which are connected to a scanned row conductor are always ignited such that the non-selected display elements of that row receive a low quiescent current through quiescent current supply circuits in the column selection circuits and the selected display elements receive a higher display current through selection circuits in the column selection circuits. The low quiescent currents reduce the ignition-voltage required for the next following row to be scanned while a high contrast is produced between the selected and non-selected display elements.

---

4 171 363

Hot-gas engine

P. T. J. Bex E

A hot-gas engine comprising an engine part which is provided with a blanket of ceramic fibres in order to provide protection against high flame temperatures of the burner. The blanket is anchored to the engine part by means of lugs which are bent around metal wires arranged between the fibres.

---

4 172 756

Method for the accelerated growth from the gaseous phase of crystals, and products obtained in this manner

L. Hollan L

A method of growing gallium arsenide crystals obtained by reaction of a first gaseous phase containing hydrogen and arsenic trichloride and a liquid gallium phase, in which the reaction forms a second gaseous phase from which gallium arsenide is deposited in such circumstances that the molar fraction of arsenic trichloride is larger than 2×10^{-2} and is preferably between 2×10^{-2} and 10^{-1}.

---

4 172 778

Ion-selective electrode

R. E. van de Leest E
L. Heijne

Ion-selective electrode for determining the concentration of one specific ion in a solution of a mixture of different ions. The active material consists of Ag_3SBr or Ag_3SI. The electrode is used for the selective determination of Br^-, I^- or Ag^+.

4 174 477

Method of and device for arc welding

W. G. Essers E
G. Jelmorini
G. W. Tichelaar

A method of plasma MIG welding in which an arc is maintained between a workpiece and a welding wire in a thermally ionized gas (auxiliary plasma) which is generated by an electric gas-discharge between two non-consumable electrodes; neither the welding wire nor the workpiece acts as a single electrode for generating the thermally ionized gas.

---

4 174 616

Insulated cylinder sleeve for a hot-gas engine

A. M. Nederlof E
G. A. A. Asselman

A hot-gas engine having a heat insulating lining of loose grains surrounding the hot working medium spaces of the cylinder and the regenerator housing.

---

4 175 010

Method of reinforcing a conductive base pattern by electroplating and device obtained by means of the method

G. J. Koel E
J. Gerkema

A method of manufacturing a device in which a conductive base pattern is formed on a substrate and is reinforced by electroplating so as to obtain a conductor pattern, characterized in that at least a surface layer of the substrate consists of an electrically insulating material on which an auxiliary layer is provided having a sheet resistance between 10^{12} and 10^{17} ohms per square, after which the base pattern is formed on the last-mentioned layer.

---

4 176 016

Forming electrically insulating layers by sputter deposition

G. J. Koel E
L. Postma

A method of making inclined sides on a metal pattern which is grown on a substrate by electroplating as a preliminary measure to the coating of the substrate with the metal pattern by an insulating layer. The main deposition of the insulating layer by sputtering is preceded by a sputter-etching step which transforms the rounded edges of the metal pattern into inclined sides.

---

4 176 336

Lacquer-encapsulated carbon film resistor

P. J. T. van den Berk E
G. Verspui

A lacquer-encapsulated carbon film resistor having a ceramic substrate coated with a carbon film, a silicon nitride layer, an organic lacquer layer and termination electrodes.

---

4 176 381

Recording and/or reproducing apparatus for a magnetic record carrier in the form of a tape, provided with a control system for the magnetic head position

E. de Niet E
A. M. A. Rijckaert

A recording and/or reproducing apparatus for a record carrier in the form of a tape, with a plurality of parallel longitudinal tracks. The apparatus comprises positioning means for controlling the position of the write/read head transverse to the direction of movement of the record carrier. The control signal for these positioning means is obtained with the aid of two optical systems, which each comprise a light-emitting element for emitting a radiation beam which is aimed at one edge of the record carrier and a light-receiving element for picking up the amount of light which shines past said record carrier, at least one of the elements of each optical system being rigidly connected to the write/read head. A detection device is coupled to the two light-receiving elements, which device derives the desired control signal from the amounts of light received by these two light-receiving elements.

---

4 178 077

Electrostatically controlled picture display device

T. S. te Velde E

A passive picture display device having a number of display elements for controlling the reflection or the transmission of light. Each display element has two parallel supporting plates of which at least one is transparent, a first electrode disposed on one of said supporting plates, a second electrode disposed on the other supporting plate facing the first electrode, a third electrode movably disposed between the first and second electrodes by electrostatic forces and an opaque liquid the color of which contrasts with the color of the side of the third electrode facing the transparent supporting plate filling the space between the two supporting plates.

---

4 178 510

Device for measuring the spatial distribution of radiation absorption in a slice of a body

W. Wagner H

In third-generation transversal tomography scanners, where the total body slice to be examined is irradiated by a flat, fan-shaped radiation beam, the probability that a detector which measures part of the radiation in the radiation beam becomes defective is very high. This failure causes artefacts in the reconstruction of the absorption distribution of the slice of the body, said artefacts having an adverse effect on the evaluation of the image by the radiologist. The invention ensures that a defective detector is identified (for example, by comparison of the measuring values supplied by each detector with the measuring values of adjacent detectors), the measuring value of a detector which has been identified as being defective being replaced by a correction value which can be formed by interpolation of measuring values obtained along neighboring paths.

---

4 178 530

Electron tube with pyrolytic graphite heating element

B. Lersmacher A
H. Lydtin
H. Seifert
J. W. A. Krol

In a thermionic cathode having a planar emissive body and a heating element of pyrolytic graphite which is provided on the side of the emissive body remote from the emissive surface of the emissive body, a uniform temperature distribution adjusts during operation throughout the overall emissive surface when the heating element is planar and the crystallographic *c*-axis of the pyrolytic graphite extends everywhere normal to the surface of the heating element facing the emissive body. As a result of this the possibility is obtained of realizing a planar, 'rapid', induction-free unipotential cathode having a substantially ideal homogeneous temperature distribution.

---

4 178 570

Micro-slot-stripline phase shifter

B. Schiek H
J. Köhler

A phase shifter having, on one surface of a substrate, a microstrip stub which is connected in parallel to a microstrip line and which is open at one end. Disposed below the stub is a slot which is short-circuited at a point below the open end of the stub, and which ends in an open circuit at the opposite end.

4 179 166

Anti-lock vehicle brake system with variable decay rate

D. Sharp R
D. R. Skoyles

An anti-lock vehicle brake system includes a source of fluid pressure, a wheel brake connected to the source of fluid pressure and an anti-lock control valve for relieving pressure at the brake. Sensing and control apparatus are provided for positioning the control valve. A restriction apparatus limits the flow of fluid pressure to the brake. A fluid flow control is provided intermediate the anti-lock control valve and the restriction apparatus for control of the restriction apparatus.

---

4 179 687

Analog-to-digital converter circuit employing iterative subtraction

R. J. van de Plassche E
E. C. Dijkmans

An analog-to-digital conversion circuit having a number of series-connected stages, each stage determining the difference between a signal current input and a reference current input, and at positive values of said difference transferring a current proportional to said difference to the succeeding stage. In each successive stage, the reference current is subtracted from the signal current until the residual current is smaller than the reference current. The number of successive stages in which this subtraction takes place can be detected to indicate the level of the input current. Each stage includes a common point for receiving the difference between the signal current and the reference current, a current path from said common point to its output for transferring a current proportional to the difference if positive, a current path from said common point to a current sink if said difference is negative; said current paths being unidirectional and biased such that only one is forward biased at one point in time. Each stage also detects the polarity of said difference to obtain a digital measure of the value of the signal current.

---

4 179 777

Method of producing a low-pressure gas discharge lamp

J. Hasker E
C. Peters

A method of producing low-pressure gas discharge lamps wherein filamenting wool is prepared into a mat and then brought into the discharge space.

---

4 180 761

Control device, particularly for controlling the emission current of an X-ray tube

R. Ochmann H

A control device for controlling the emission current of an X-ray tube comprising a control circuit which receives a digital reference value signal and an amplifier whose gain is controlled by a portion of said digital reference value signal so as to provide a comparatively high accuracy for the emission current adjustment with a minimum number of binary positions.

---

4 180 785

Modulator for forming the product of two input signals

J. O. Voorman E

A modulator suitable for amplitude modulation includes first and second converters for converting the positive and negative excursions of first and second input signals into four respective signals of one polarity which are passed through a multiplying device and a difference producer to generate an output signal containing only the product terms of the first and second input signals.

4 181 155

Pressure energy accumulator

G. A. Wesselink E
A. J. J. Franken
H. F. G. Smulders

A pressure energy accumulator having a mixture consisting of CF_4 or C_2F_6 and N_2, H_2, or He, or a combination thereof, as the gaseous working medium.

---

4 182 973

Electric discharge tube

G. A. H. M. Vrijssen E
F. C. M. de Haas

Disclosed is a bayonet catch for directly securing an electrode in a glass envelope which comprises projections extending inwardly from the envelope and radially extending portions on the electrode which are clamped between the projections upon rotation of the electrode. The material of the envelope and the electrode have substantially equal coefficients of expansion to minimize relative movement between the envelope and the electrode with changes in temperature.

---

4 183 037

Semiconductor device

C. J. P. F. Le Can E
C. M. Hart
H. E. J. Wulms

The invention relates to a semiconductor device in which a crossing connection is realized by using parts of a layer of refractory conductive material already present for masking as a part of a current conductor separated from a crossing conductor by an insulation layer. The mask of refractory material may also define the regions in which switching transistors are realized. The invention results in important advantages, in connection with density and crossing connections, in particular in I^2L-circuits.

---

4 185 201

X-ray detector

A. L. N. Stevels E

The X-ray detection element of an X-ray detector is composed of a fluorescent material having a short afterglow and a high sensitivity to the X-radiation to be detected. For the fluorescent material use is notably made of cerium-activated phosphors such as yttrium phosphors, lanthanum phosphors or lutetium phosphors.

---

4 185 236

Current stabilizer

R. J. van de Plassche E
A. C. M. Schepens

A current stabilizing arrangement with a first and a second current circuit in which currents with a fixed ratio relative to each other are sustained. These currents respectively flow through a first semiconductor junction and a second semiconductor junction, a resistor being alternately connected in series with the first and second semiconductor junction. The currents flowing in the first and the second current circuit are interchanged in synchronism therewith.

---

4 185 795

Magnetic tape drive with friction compensation

A. Walraven E

In a magnetic tape drive arrangement having a tension-controlled loop, with the tension in the part of the magnetic tape upstream from the loop being maintained constant as far as possible by suitable means in a conventional manner, compensation of variations of the tension of the magnetic tape in the loop as the result of variations in the coefficients of friction of the magnetic heads due to changes in temperature and humidity, is provided by an element which frictionally engages the magnetic tape upstream from the closed loop, coefficient of friction and angle of contact of the element being selected in accordance with a specific relationship.

PHILIPS TECHNICAL REVIEW

VOLUME 39, 1980, No. 3/4

A microwave-induced plasma as an excitation source for atomic emission spectrometry

C. I. M. Beenakker, P. W. J. M. Boumans and P. J. Rommers

A microwave-induced plasma (MIP) is one of the excitation sources that have been investigated during the last ten or twenty years for their usefulness in elemental analysis by means of atomic emission spectrometry. With the resonant cavities used until recently it was necessary to take special precautionary measures for generating the plasma and for maintaining it while introducing a sample for analysis. The authors have designed a resonant cavity in which the microwave energy is transferred to the plasma more efficiently, resulting in an MIP that is more easily generated and more stable, and can be used for a wider range of analytical applications.

Introduction

In atomic emission spectrometry (AES) a hot partly ionized gas (plasma) is used as an excitation source. A sample introduced into this source gives an emission spectrum with characteristic lines for the various elements, so that with a suitable optical detection system it is possible to carry out sequential or simultaneous multi-element analyses. There are various ways in which a plasma can be generated. An earlier article in this journal described how an argon plasma is generated and kept stable by means of induction heating (50 MHz) [1]. In this article we shall be concerned with a microwave-induced plasma (MIP). To produce the plasma a microwave generator operating at a frequency of 2450 MHz is connected by a coaxial cable to a resonant cavity in which there is a discharge tube containing a gas. A plasma is generated when a number of conditions relating to the microwave power, the dimensions of the cavity and the nature and pressure of the gas in the discharge tube are satisfied. A sample introduced into a plasma becomes dissociated, and the atomic or molecular fragments are then excited. From the resultant optical emission spectrum the elements present in the sample and their concentrations can be derived.

The preferred gases for the plasma are helium and argon, since they give no molecular background spectra themselves. The use of helium has the advantage that elements that have a high excitation energy can also be detected by means of line emission spectra. When argon is used there are only the band emission spectra of the diatomic molecules of such elements as nitrogen, oxygen, fluorine, chlorine and bromine. These spectra can sometimes be difficult to identify [2], and there may be spectral overlap between the different bands. Also, the band intensity does not always increase linearly with the concentration of the element [3]. An initial disadvantage of using helium rather than argon was that with the conventional resonant cavities a stable MIP could only be generated at reduced pressure (about 10^4 Pa [4,5]). This requires the use of a vacuum pump and in addition the connecting tubes must be thin to obtain the required pressure drop between the plasma and the position where the sample is introduced.

Dr C. I. M. Beenakker, Dr P. W. J. M. Boumans and Ir P. J. Rommers are with Philips Research Laboratories, Eindhoven.

[1] P. W. J. M. Boumans, F. J. de Boer and J. W. de Ruiter, Philips tech. Rev. 33, 50, 1973.
A recent survey of plasma sources for spectrochemical analyses is given in the article by P. Tschöpel in: G. Svehla (ed.), Comprehensive analytical chemistry, Vol. IX, Elsevier, Amsterdam 1979 (chapter 3).

[2] A. J. McCormack, S. C. Tong and W. D. Cooke, Anal. Chem. 37, 1470, 1965.

[3] R. M. Dagnall, T. S. West and P. Whitehead, Anal. Chem. 44, 2074, 1972.

[4] F. C. Fehsenfeld, K. M. Evenson and H. P. Broida, Rev. sci. Instr. 36, 294, 1965.

[5] S. Greenfield, H. McD. McGeachin and P. B. Smith, Talanta 22, 553, 1975.

An MIP is very suitable for applications where it is combined with a gas chromatograph [2]. It can provide an element-specific analysis of the emergent gas. The carrier gas of the chromatograph is also used as the plasma gas. In our investigation into the usefulness of a helium MIP as a detector for gas chromatography we found that the necessity to work at reduced pressure was due to an inadequate transfer of the microwave energy through the resonant cavity to the plasma [6]. The cavity that we have designed gives a much more efficient energy transfer, and a stable MIP can be obtained even in helium at atmospheric pressure [6,7].

A plasma generated with the new cavity can also be used as an excitation source for elemental analyses of solutions [5,8]. The solutions are 'nebulized' to form an aerosol, which is injected into the plasma. The nebulization is best with argon; helium is too light. Previously it was not very easy to keep an argon MIP stable at atmospheric pressure when a wet aerosol was introduced, but this can be done with the new resonant cavity [9,10].

A disadvantage of an argon MIP for analysing solutions is that the plasma does not fill the discharge tube homogeneously. Because of this a large proportion of the aerosol may pass through the discharge tube without sufficient excitation of the elements contained in it. A homogeneous filling can however be achieved with a helium plasma. The use of a helium MIP also has the advantage that the analysis can be extended to include such elements as nitrogen, oxygen and the halogens. For the analysis of solutions we have therefore investigated the combination of a helium MIP with a method of sample introduction used in atomic absorption spectrometry, where the solutions are not nebulized but electrothermally evaporated and atomized. In electrothermal atomization a measured amount of sample solution is placed upon a suitable substrate situated in the gas stream in front of the plasma. After evaporation of the solvent, the substrate is electrically heated for a few seconds to a temperature between 2000 and 3000 °C. The evaporating and partly atomized substance is then carried along with the helium gas stream to the plasma, where further atomization takes place. The sample is not continuously introduced here, as in the combination of an MIP and a nebulizer, but in discrete amounts; hence the term 'discrete sampling'. These amounts can be extremely small (a few μl), so that the method is also very suitable for samples that are only available in very small quantities.

In this article we shall first describe the equipment used for generating and sustaining an MIP, with particular emphasis on the new cavity. We shall then discuss the excitation mechanisms that can occur in an MIP. Next we shall consider the likely range of analytical applications of an MIP as an AES excitation source. Here we shall pay particular attention to the combination of an MIP with a gas chromatograph, a nebulizer, and an electrothermal atomizer. Finally, we shall try to assess the current status of the MIP-AES method in analytical spectrometry.

The resonant cavity

The main components of the equipment for generating an MIP are a microwave generator, a resonant cavity and a discharge tube. An MIP is produced as soon as the magnitude of the electric component of the electromagnetic field pattern in the resonant cavity exceeds a critical minimum value, which is partly determined by the nature and the pressure of the plasma gas. With standard equipment using a microwave frequency of 2450 MHz and a microwave power of less than 200 W this value cannot be reached with a conventional resonant cavity [4] for helium and wet argon at atmospheric pressure. A higher electric field-strength can be obtained by designing the cavity to meet three conditions: the resonant frequency of the (filled) cavity must be accurately tuned to the microwave frequency, the volume of the cavity must be as small as possible to give the highest energy density for a given power, and the discharge tube must be located at the position where the electric field-strength is at a maximum.

The dimensions of a cavity resonator are usually completely determined by the resonant frequency and by the desired configuration of the field pattern (i.e. the 'mode'). However, this is not so for the transverse magnetic modes TM_{mnp} with $p=0$ in a cylindrical cavity resonator, where m, n and p relate to the number of maxima of the field pattern in the angular, radial and longitudinal directions respectively [11]. This is because the resonant frequency for the TM_{mn0} modes is independent of the dimension in the longitudinal direction. Thus, in a cavity operating in a TM_{mn0} mode the electric field can be increased, for constant power, by reducing the height of the cylinder.

The cavity diameter for resonance in a TM_{mn0} mode is smallest when $m=0$ and $n=1$. At a resonant frequency of 2450 MHz the calculated value of the diameter is 93.7 mm. Owing to the presence of a dielectric — in the form of a discharge tube — the actual resonant frequency has a somewhat lower value. The desired diameter then has to be calculated for a resonant frequency slightly higher than the generator frequency. This means that the actual diam-

eter should be made somewhat smaller than the calculated one. In the first experimental version small changes in the resonant frequency due to the presence of the plasma and differences between discharge tubes, etc., were corrected by means of tuning screws in the cylindrical wall or the top of the cavity. A 'double-stub tuner' is used for adjusting the impedance of the cavity to obtain the best possible power transfer ('matching') [9].

The diagram of *fig. 1* shows how the electric field-strength varies as a function of the radial distance in a TM_{010} mode. The axial electric field has a maximum in the centre of the cavity. The discharge tube is therefore mounted along the axis in the centre of the cavity.

Copper is used as the material for the resonant cavity because of its high electrical conductivity. Silica is generally used for the discharge tube because it has a relatively high melting point and its dielectric losses are low. Since the optical transmission of silica is high, emission signals can in principle be detected in all directions. The use of silica can occasionally lead to problems, however, as in the determination of fluorine and silicon, and also if water is present in the plasma gas. This can cause erosion of the silica and hence errors in the detected signals. In such cases

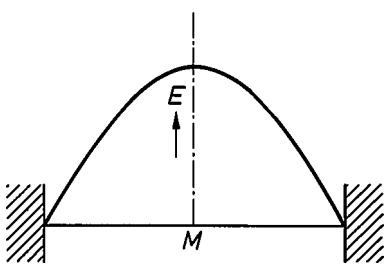


Fig. 1. Electric field-strength E in a TM_{010} cavity. The field-strength has a maximum at the centre M.

other electrically non-conducting materials with a high melting point can be used, such as aluminium oxide or boron nitride. Owing to the low optical transmission of these materials the emission signals can only be detected along the axis.

With the TM_{010} cavity described here it has been found possible to generate a stable MIP at atmospheric pressure not only in dry argon but also in wet argon and in helium [6,7,9,10]. The first results obtained with this cavity have prompted a number of investigators elsewhere to make a closer study of its capabilities [12–17]. Their results provide a general confirmation of our ideas on the suitability of this cavity, especially for producing plasmas at atmospheric pressure in helium. They have also prompted a few modifications that make the cavity easier to use. For instance the cavity is now tuned to the correct frequency by means of a sliding silica rod and the impedances are matched by a variable coupling probe [15]. As a result the diameter of the cavity is less critical, the height has been reduced and a double-stub tuner is no longer necessary.

Fig. 2 shows a diagram of the cross-section of the new resonant cavity now in use. *Fig. 3* shows photographs of argon and helium plasmas obtained with it. The differences between the two plasmas mentioned in the introduction can clearly be seen. The argon plasma extends a long way beyond the cavity and is generally split into two or more branches that tend to adhere to the wall of the discharge tube. The helium plasma, on the other hand, is located entirely within the cavity and fills the discharge tube homogeneously. During our experiments we checked the stability of the plasma regularly by measuring the signal from the atomic helium emission at 471.3 nm. The intensity of this signal was found to fluctuate by no more than 3% during working periods of eight hours after a warming-up period of half an hour. In a period of a few months the fluctuation was never more than about 10%.

Excitation mechanisms in an MIP

Our investigations have shown that the advantages of working with helium at reduced pressure instead of with argon at atmospheric pressure are also obtained with helium at atmospheric pressure. The emission spectra we have obtained with the new resonant cavity in atmospheric helium contain lines that do not appear in the spectra obtained in atmospheric argon. Some elements that do not give characteristic line emission in argon do so in helium. To explain these differences it is necessary to know something about

[6] C. I. M. Beenakker, Spectrochim. Acta **31B**, 483, 1976.
[7] C. I. M. Beenakker, Spectrochim. Acta **32B**, 173, 1977.
[8] R. K. Skogerboe and G. N. Coleman, Appl. Spectrosc. **30**, 504, 1976.
[9] C. I. M. Beenakker and P. W. J. M. Boumans, Spectrochim. Acta **33B**, 53, 1978.
[10] C. I. M. Beenakker, B. Bosman and P. W. J. M. Boumans, Spectrochim. Acta **33B**, 373, 1978.
[11] C. P. Poole Jr., Electron spin resonance, Interscience, New York 1967.
[12] J. P. J. van Dalen, P. A. de Lezenne Coulander and L. de Galan, Anal. chim. Acta **94**, 1, 1977.
[13] A. T. Zander and G. M. Hieftje, Anal. Chem. **50**, 1257, 1978.
[14] B. D. Quimby, P. C. Uden and R. M. Barnes, Anal. Chem. **50**, 2112, 1978.
[15] J. P. J. van Dalen, P. A. de Lezenne Coulander and L. de Galan, Spectrochim. Acta **33B**, 545, 1978.
[16] K. J. Mulligan, M. H. Hahn, J. A. Caruso and F. L. Fricke, Anal. Chem. **51**, 1935, 1979.
[17] B. D. Quimby, M. F. Delaney, P. C. Uden and R. M. Barnes, Anal. Chem. **51**, 875, 1979.

the nature and the state of excitation of the atomic and molecular particles in the plasma, the excitation processes in which they may be involved, and the excitation energies of the atoms and ions of the elements to be determined. In our discussion of excitation mechanisms we shall confine ourselves to the strong emission lines between 190 and 800 nm that have been observed for the elements hydrogen, carbon, nitrogen oxygen, fluorine, phosphorus, sulphur, chlorine, bromine and iodine. We have chosen these elements because the excitation energies for the various emission lines cover a wide range of values. *Table I* gives the wavelengths of the various emission lines with the associated excitation energies, and also indicates whether they are also found in an argon plasma. The emission lines originate both from neutral atoms and from ions.

Possible excitation processes

In describing the possible excitation processes in an MIP we have to take account of metastable atoms and molecules, ions and electrons. In the equations of the reactions that can occur, A denotes the atoms and molecules of the plasma gas and X denotes the element to be determined (analyte). The superscripts have the following significance: m metastable, $*$ excited, $+$ singly ionized, $++$ doubly ionized.

Metastable particles A^m can take part in various processes. The principal ones are:

$$A^m + X \rightarrow A + X^+ + e, \quad (1)$$

$$A^m + X^+ \rightarrow A + X^{++} + e, \quad (2)$$

$$A^m + X \rightarrow A + X^*, \quad (3)$$

$$A^m + X^+ \rightarrow A + X^{+*}. \quad (4)$$

Processes (1) and (2) belong to the well-known Penning ionization reactions in gases [18]. These may occur if the energy of the metastable particles before the collision is greater than the first or the second ionization energy (ionization potential) of X. If the thermal energy is small, as is the case with an MIP in helium or argon, the following condition applies for the occurrence of process (1):

$$E_m(A) \gtrsim E_{ion}(X), \quad (5)$$

where $E_m(A)$ is the metastable energy of A and $E_{ion}(X)$ is the ionization energy of X. A similar condition applies to the occurrence of process (2). By comparing the values that $E_m(A)$ can assume (*Table II*) with the values of the first and second ionization energies (*Table III*) we can therefore identify the ionization reactions that can occur. In processes (3) and (4) there is a direct excitation by the metastable particles.

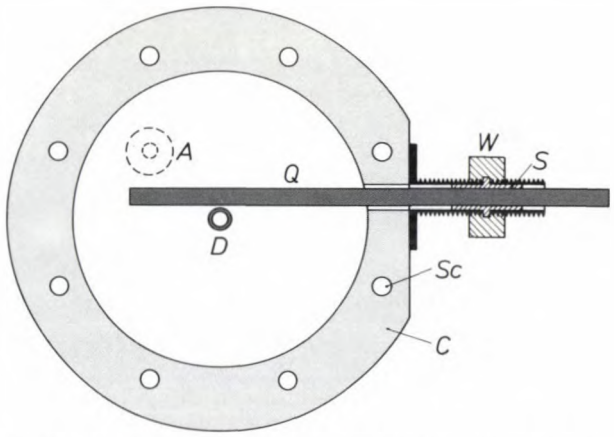


Fig. 2. Diagram of the cross-section of the new resonant cavity (half full scale) for a microwave-induced plasma (MIP). *C* cylinder wall. *Sc* tapped holes for attaching copper top plates. *Q* silica tuning rod. *S* slide clamp. *W* nut for fine adjustment of penetration. *A* position of variable coupling probe on the top plate. *D* plasma discharge tube, perpendicular to plane of drawing. Height of cavity 6 mm.

Table I. Wavelength λ with associated excitation energy E_{exc} of the emission lines obtained with an MIP for a number of elements. The values marked with an asterisk relate to emission lines that are observed both with a helium and with an argon MIP; the other emission lines were found only with a helium MIP. I the neutral atom. II the singly charged ion.

Element	λ (nm)	E_{exc} (eV)	Element	λ (nm)	E_{exc} (eV)
H	656.3	12.09	S(II)	545.4	15.88
	486.1	12.75		543.3	15.83
	434.0	13.06		560.6	15.88
	410.2	13.22		564.0	15.83
C(I)	193.1	7.68	Cl(II)	479.5	15.89
	247.9*	7.68		481.0	15.88
				481.9	15.88
N(I)	746.8	11.99		542.3	15.89
	744.2	11.99			
			Br(II)	467.9	
O(I)	777.2	10.74		470.5	14.28
	777.4	10.74		478.6	14.23
				481.7	14.22
F(I)	685.6	14.50		518.2	14.60
	690.3	14.53			
			I(I)	206.2*	6.92
P(I)	253.6*	7.18			
	255.3	7.14	I(II)	466.6	15.10
	253.4	7.18		516.1	12.45
	255.5*	7.18		533.8	14.30
	213.6	7.18		534.5	14.30
	214.9	7.14		540.7	13.91
				543.6	12.72
P(II)	458.8	15.44		546.5	12.32
	460.2	15.48		549.7	12.30
				562.6	12.72
S(I)	190.0	6.50			
	191.5	6.50			
	216.9*	6.86			
	182.0*	6.86			

Table II. Energy of the lowest metastable state of atomic helium [19], neutral and ionized diatomic helium [7] and atomic argon [19].

He	He_2	He_2^+	Ar
19.73 eV	13.3-15.9 eV	18.3-20.5 eV	11.50 eV

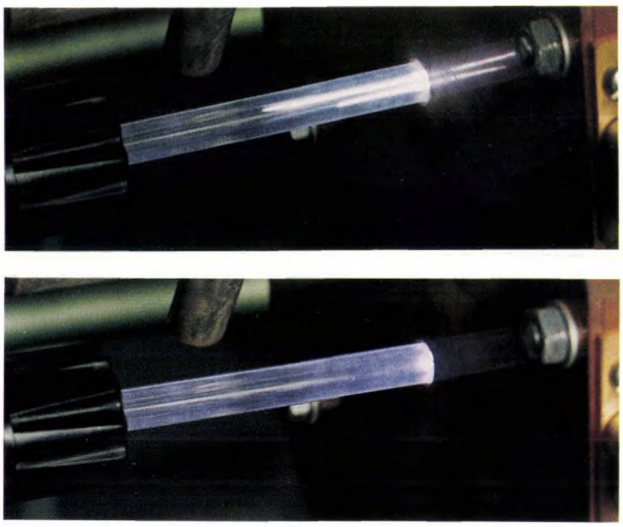


Fig. 3. Photographs of an argon plasma (*above*) and of a helium plasma (*below*) obtained with the new cavity. The helium plasma is more concentrated in the cavity than the argon plasma and is much more homogeneous.

For the occurrence of each of these processes a more stringent condition applies. For process (3) this condition is:

$$E_m(A) \approx E_{exc}(X), \tag{6}$$

where $E_{exc}(X)$ is the excitation energy of X.

Ions A^+ may be involved in excitations of the type [18]

$$A^+ + X \rightarrow A + X^{+*}. \tag{7}$$

The occurrence of such a reaction is subject to the condition:

$$E_{ion}(A) \approx E_{ion}(X) + E_{exc}(X^+). \tag{8}$$

Electrons with a relatively low energy are particularly effective in a plasma in bringing about recombinations:

$$e + A + X^+ \rightarrow A + X^*. \tag{9}$$

Table III. The first and second ionization energies E^1_{ion} and E^2_{ion} of the elements given in Table I and of helium and argon [20].

Element	E^1_{ion} (eV)	E^2_{ion} (eV)
H	13.60	
He	24.59	54.42
C	11.26	24.38
N	14.53	19.60
O	13.62	35.12
F	17.42	34.97
P	10.49	19.73
S	10.36	23.33
Cl	12.97	23.81
Ar	15.76	27.63
Br	11.81	21.8
I	10.45	19.23

Fast electrons, on the other hand, maintain the plasma by reactions of the type

$$e + A \rightarrow A^+ + 2e. \tag{10}$$

In addition, fast electrons may also be directly involved in excitation processes:

$$e + X \rightarrow X^* + e, \tag{11}$$

$$e + X^+ \rightarrow X^{+*} + e. \tag{12}$$

We shall now show which mechanisms give rise to line emission in a helium or argon plasma.

Excitation in a helium plasma

In a helium plasma all the elements give line emission. Some give atomic lines only, whereas others give ionic lines as well. If we look at the excitation energies of the ionic lines of phosphorus, sulphur, chlorine, bromine and iodine, we see that these energies remain limited to a fairly narrow range (12.3-15.9 eV), even though these ions are known to have many energy levels situated both below and above this range and these can also give rise to spectral transitions in the wavelength range between 190 and 800 nm [19,21]. It can be seen from this that direct excitation by electrons (eq. 12) is not very important. The continuous energy spectrum of electrons cannot be made to correspond to the observed limitation of the number of populated energy levels. Moreover, the relative intensities in the spectra for electronic excitation would have to depend strongly on the pressure, because the energy of the electrons is highly pressure-dependent.

Reactions of the type of eq. (7) cannot come into the picture either. It can be shown from the excitation energies for the emission lines of the singly charged ions and from the ionization energies of the associated atoms and of helium that the condition for the occurrence of such reactions (eq. (8)) is not satisfied.

The excitation energies of the observed emission lines of the ions, apart from a few of the lines of the iodine ion, fall neatly in the range (13.3-15.9 eV) covered by a metastable helium molecule. For all

[18] M. J. Druyvesteyn and F. M. Penning, Rev. mod. Phys. **12**, 87, 1940.

[19] C. E. Moore, Revised multiplet table of astrophysical interest, NSRDS-NBS 40, U.S. Government Printing Office, Washington 1972.

[20] J. L. Franklin, J. G. Dillard, H. M. Rosenstock, J. T. Herron, K. Draxl and F. H. Field, Ionization potentials, appearance potentials, and heats of formation of gaseous positive ions, NSRDS-NBS 26, U.S. Government Printing Office, Washington 1969.
C. E. Moore, Ionization potentials and ionization limits derived from the analyses of optical spectra, NSRDS-NBS 34, U.S. Government Printing Office, Washington 1970.

[21] A. R. Striganov and N. S. Sventitskii, Tables of spectral lines of neutral and ionized atoms, IFI/Plenum, New York 1968.

these lines the following condition is therefore satisfied:

$$E_m(He_2) \approx E_{exc}(X^+). \quad (13)$$

This strongly suggests that the emission lines of the singly charged ions here are produced through reactions of the type of eq. (4):

$$He_2^m + X^+ \rightarrow 2He + X^{+*}. \quad (14)$$

The X^+ ions can be formed by collisions with electrons, but they are mainly formed by Penning ionization reactions, because the energy of metastable helium is high enough for all these elements to be ionized.

Another process is evidently responsible for the emission lines of the singly charged iodine ion, which have an excitation energy below 13.3 eV. Since the second ionization energy of the iodine atom is smaller than the energy of a metastable helium atom, the reaction of eq. (2) may also occur for this element:

$$He^m + I^+ \rightarrow He + I^{++} + e. \quad (15)$$

The doubly charged iodine ion can then become an excited singly charged iodine ion by recombination with an electron.

It might seem from the above that the process of eq. (14) could also occur for the singly charged ions of carbon, nitrogen, oxygen, and fluorine. A glance at the energy diagrams of these ions [21] makes it clear, however, that there are no energy levels that could be reached by reaction with metastable helium and could give emission in the wavelength range from 190 to 800 nm. Nor can reactions of the type of eq. (2) occur here, because the second ionization energy of these elements is too high compared with the energy of a metastable helium atom. Excitation of the singly charged ions with fast electrons (eq. (12)) is not very easy here either, because the excitation energies are fairly high. The only remaining possibility for these ions is therefore a recombination with low-energy electrons (eq. (9)), leading to the formation of excited neutral atoms.

Excitation in an argon plasma

With an argon MIP no ionic lines are observed for the elements discussed here, but only atomic lines, and then only for carbon, phosphorus, sulphur and iodine (Table I). These findings can be explained if we assume that the ionization in an argon plasma takes place mainly by way of Penning processes of the type of eq. (1). The energy of a metastable argon atom is insufficient for the ionization of nitrogen, oxygen, fluorine, chlorine and bromine (Tables II and III). The only elements whose ionization energies are low enough for the formation of the singly charged ions by interaction with metastable argon atoms are carbon, phosphorus, sulphur and iodine. However, the energy levels of these ions from which emission can occur in the wavelength range from 190 to 800 nm are all above the energy levels of metastable atomic argon. This implies that these ions cannot be excited by metastable argon atoms, and certainly not by metastable argon molecules. These ions can only recombine with electrons (eq. (9)), producing excited neutral atoms that give the emission lines mentioned in Table I.

The formation of the line emission spectra in a helium MIP and an argon MIP can thus be described with a relatively simple reaction pattern in which metastable atoms and molecules of the plasma gas play an important part. The knowledge obtained in this way enables us to explain why a helium MIP at atmospheric pressure gives smaller minimum detectable concentrations than at lower pressure. This is because an increase in pressure gives an increased concentration of metastable helium [22], and consequently the emission lines of the singly charged ions of phosphorus, sulphur, chlorine, bromine and iodine become stronger. At the same time, owing to the decrease in the electron energy, there are more recombinations of the type of eq. (9), so that the atomic emission lines become stronger as well.

Analytical capabilities of an MIP

Combination with a gas chromatograph

Fig. 4 is a diagram showing an MIP at atmospheric pressure combined with a gas chromatograph. Helium is the best gas for the plasma and is therefore the preferred carrier gas for the chromatograph.

In our investigation into the usefulness of an atmospheric helium MIP as an element-selective de-

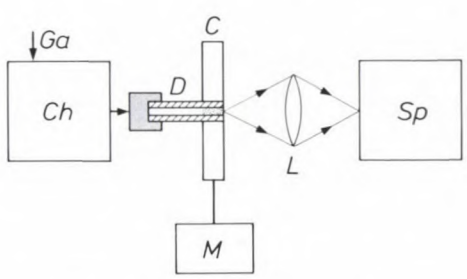


Fig. 4. Diagram showing the combination of an MIP at atmospheric pressure and a gas chromatograph *Ch*. Helium or argon is used as the carrier gas Ga of the chromatograph, and also serves as the feed gas for the plasma. Gaseous components of the material to be analysed are transported from the gas chromatograph to the discharge tube *D*, which is partly located in the cavity *C*. *M* microwave generator. The light emission from the plasma is detected by the spectrometer *Sp* via the silica lens *L*.

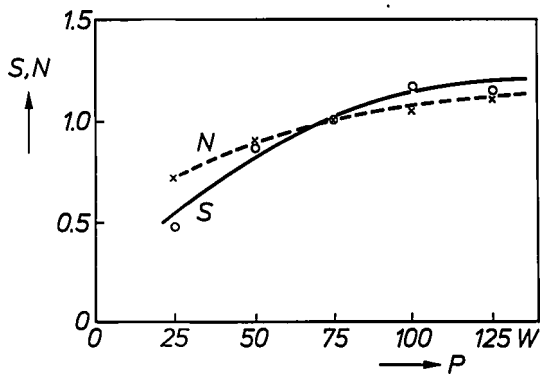


Fig. 5. Net line intensity S of carbon emission at 193.1 nm and background noise N in arbitrary units, as a function of the microwave power P. The values of S and N at 75 W have been normalized to 1. The ratio S/N is not very dependent on P.

tector in a gas chromatograph, we started with a large number of liquid samples of known composition and volume. We evaporated these samples and diluted them with helium gas to the desired concentrations. The gas mixtures thus obtained were then fed to a plasma, and the line emission produced in the plasma was detected with the aid of a monochromator and a photomultiplier.

The intensities of the emission lines with respect to the background noise depend to some extent on the power of the 2450 MHz generator. This is illustrated in *fig. 5* for the emission line of carbon at 193.1 nm. We have obtained similar results for the emission lines of other elements. The power was generally 75 W in our experiments. In this case the signal-to-noise ratio, which determines the minimum detectable concentration, does not then differ very much from the maximum value (fig. 5) and it is not necessary to cool the resonant cavity. We also found that the signal from the emission lines and the signal from the background are virtually independent of the gas flow. This characterizes the MIP as a concentration-sensitive detector. In conjunction with a gas chromatograph stronger emission lines are therefore obtained for the same absolute amount of substance and a lower gas flow as a result of the higher concentration of the elements to be determined in the plasma [14,17].

In general the usefulness of an analytical method for determining the concentration of a particular element is largely determined by the sensitivity, the minimum detectable concentration, the selectivity and the linear range.

The relative sensitivity S is defined as the relative intensity of the measuring signal per unit concentration of the particular element. *Table IV* gives values of S for different emission lines obtained in an MIP analysis of a number of mainly organic compounds. The relative sensitivity for the compounds investigated is found to be not very dependent on the chemical environment of the elements. It therefore seems to be possible to use a helium MIP at atmospheric pressure to determine relative concentrations for different elements in the same molecule from the relative intensities of the various emission lines [12].

Table IV. Relative sensitivity for a number of analysis lines of different elements measured with a combination of an MIP and a gas chromatograph for various compounds. The values for chloroform are normalized at 1.00.

Analysis line / Compound	C 193.1 nm	H 486.1 nm	Cl 479.5 nm	S 545.4 nm
Hexane	0.97	0.91		
Chloroform	1.00	1.00	1.00	
Monochlorobenzene	0.95	0.95	0.95	
Tetrachloromethane	1.07		1.06	
Methylene bromide	0.99	0.93		
Methyl iodide	1.05	0.95		
Carbon disulphide	1.01			1.00
Thiophene	0.73	0.97		0.92

For the minimum detectable concentration c_{min} we use the following definition:

$$c_{min} = 2\,\sigma_B/S, \qquad (16)$$

where σ_B is the standard deviation of the background signal.

We define the selectivity *Sel* for a particular element as the ratio of the carbon concentration and the concentration of the element that both give an emission signal of the same strength. Hexane was used as a reference sample in our determinations.

Table V gives the values of c_{min}, *Sel* and the linear range for a number of emission lines of hydrogen, carbon, sulphur, chlorine, bromine and iodine. The values of c_{min} obtained are one to two orders of magnitude better than those obtained with a helium

Table V. Detection limit c_{min}, selectivity *Sel* and linear range, measured with a combination of an MIP and a gas chromatograph for a number of analysis lines.

Analysis line	c_{min} (mol/l)	*Sel*	Linear range
C 193.1 nm	3×10^{-11}		$\approx 3\times10^{3}$
247.9 nm	4×10^{-11}		$\approx 3\times10^{3}$
H 486.1 nm	2×10^{-9}		$>2\times10^{3}$
Cl 479.5 nm	2×10^{-10}	200	$>2\times10^{4}$
481.0 nm	3×10^{-10}	220	$>2\times10^{4}$
Br 470.5 nm	6×10^{-11}	220	$>3\times10^{4}$
478.5 nm	1×10^{-10}	220	$>2\times10^{4}$
I 516.1 nm	2×10^{-11}	130	$>2\times10^{4}$
206.2 nm	3×10^{-11}	250	$>4\times10^{3}$
S 545.4 nm	8×10^{-10}	200	$>5\times10^{3}$

[22] P. M. Houpt, Anal. chim. Acta 86, 129, 1976.

MIP at low pressure [23,24] and with an argon MIP at atmospheric pressure [1]. The selectivity can be improved without loss of detection capability by using an optical system of higher resolution [14], or by applying wavelength modulation [22]. For most elements the linear range is reasonably large.

Combination with a nebulizer

With the new resonant cavity the MIP can now be connected to a pneumatic nebulizer directly (i.e. without a drying stage); see *fig. 6*. In choosing the power of the microwave generator and the gas flow we paid particular attention to the stability of the plasma, the detection limit and the influence of other substances on the intensities of the emission signals ('interferences'). The results obtained indicated that we should use a power of 150 W and a gas flow of 1.4 l/min. We used aluminium oxide for the discharge tube, since this material, unlike silica, is not significantly eroded by the water vapour in the plasma.

Table VI gives the minimum detectable concentrations for a number of metals; the values were obtained with an argon MIP excitation source connected to a nebulizer. These values are substantially better than those obtained using conventional resonant cavities [25].

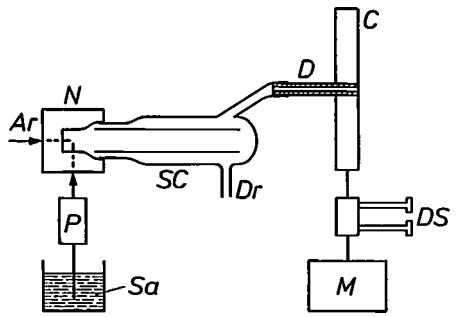


Fig. 6. Diagram of the combination of an argon MIP and a nebulizer. A liquid sample *Sa* enters a pneumatic nebulizer *N* through a peristaltic pump *P*. The aerosol formed by nebulization with argon gas enters the discharge tube *D* through a spray chamber *SC*. The experiments with the nebulizer were performed with a TM_{010} cavity in the original arrangement, with a double-stub tuner *DS* between the microwave generator *M* and the cavity *C* to give the best impedance match. *Dr* drain.

To assess the extent of interelement interferences, and in particular the interference due to a substance (the 'matrix') present in high concentration in the sample, we define the 'matrix effect' X as the intensity ratio for the analysis signal plus matrix and the analysis signal without matrix. The values for X given in Table VI relate mainly to the use of potassium chloride (2 mg/ml) as matrix. For some elements the matrix effect is found to be fairly strong. Its magnitude depends on the concentration of the matrix. *Fig. 7* shows the influence of the potassium-chloride concentration on the matrix effect for the emission lines of strontium and manganese. In general a matrix

Table VI. Wavelength λ of the analysis line, detection limit c_{min} and matrix effect X, determined with a combination of an MIP and a nebulizer for a number of metallic elements. I the neutral atom, II the singly charged ion. The detection limit of aluminium is dictated by the presence of a discharge tube of aluminium oxide. The values of X marked with an asterisk were determined with 2 mg/ml KNO_3, the others with 2 mg/ml of KCl as matrix.

Element	λ (nm)	c_{min} (µg/ml)	X
Ag(I)	338.3	0.006	1.10*
Al(I)	396.2	0.4	0.55
Co(I)	345.4	0.15	0.95
Co(I)	240.7	0.25	0.80
Cr(I)	425.4	0.15	1.10
Cu(I)	327.4	0.009	0.70
Fe(I)	372.0	0.4	0.65
Fe(II)	259.9	0.2	0.70
Ga(I)	417.2	0.01	0.85
Li(I)	670.8	0.001	2.3
Li(I)	610.4	0.01	2.2
Mg(I)	285.2	0.005	0.70
Mg(II)	279.6	0.006	1.60
Mn(I)	403.1	0.05	1.00
Mn(II)	257.6	0.03	0.80
Ni(I)	341.5	0.08	1.00
Ni(I)	352.5	0.2	0.95
Pb(I)	405.8	0.1	1.05*
Sb(I)	259.8	0.4	0.35
Sr(I)	460.8	0.01	1.30
Sr(II)	407.8	0.005	1.50
Ti(I)	498.2	0.4	0.90
Ti(II)	334.9	0.6	0.50

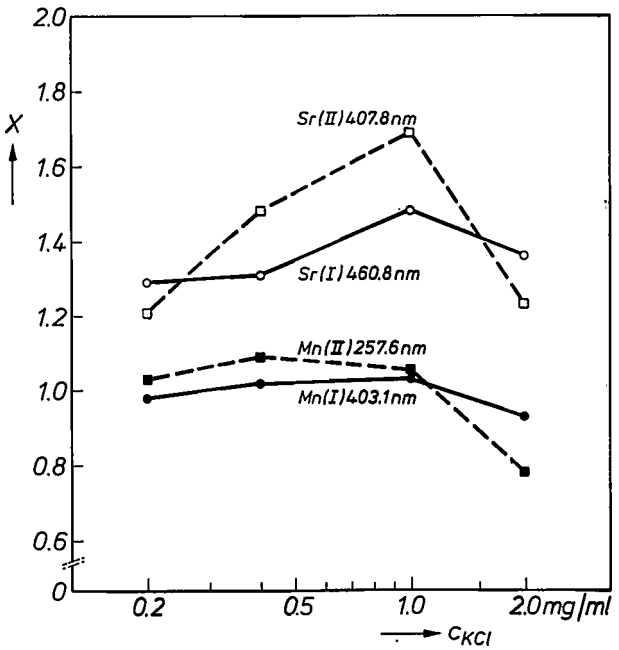


Fig. 7. The matrix effect X for emission of manganese and strontium, as a function of the KCl concentration c_{KCl}.

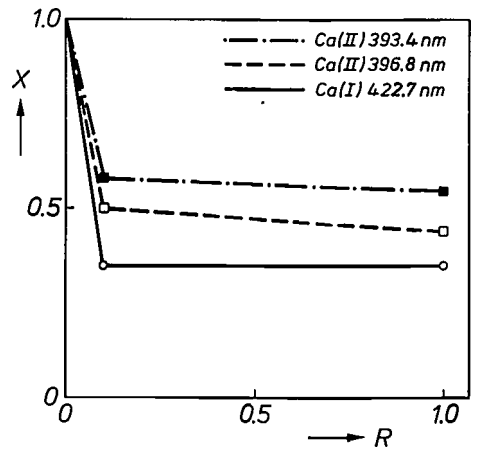


Fig. 8. The matrix effect X for the emission lines of calcium, as a function of the ratio R of the phosphate and calcium concentrations.

Table VII. Wavelength λ of the analysis line and detection limit c_{min} for a number of metallic elements, determined with an MIP, CMP or ICP in practically the same spectrometric conditions and using the same injection rates.

Element	λ (nm)	c_{min} (µg/ml)		
		MIP	CMP	ICP
Al(I)	396.2	0.4	0.3	0.02
Fe(I)	372.0	0.4	0.15	0.05
Ga(I)	417.2	0.01	0.15	0.03
Li(I)	610.4	0.01	0.7	0.05
Mg(I)	285.2	0.005	0.03	0.004
Ni(I)	352.5	0.2	0.1	0.07
Pb(I)	405.8	0.1	0.8	0.3
Ti(I)	498.2	0.4	0.3	0.9
Ti(II)	334.9	0.6	1.5	0.003

effect in a system incorporating a nebulizer and an MIP is caused by changes in the nebulization process, the atomization in the discharge tube, and the excitation conditions in the plasma, such as the electron temperature, the electron density and the spatial distributions of these quantities. Changes in the nebulization process do not generally need to be taken into account, in view of the relatively small effects found with the same nebulization system, when it is combined with an inductively coupled plasma (ICP) [26]. Changes in the atomization, on the other hand, can be important, as appears from the marked effect of the addition of a phosphate on the emission lines of calcium; see *fig 8*. These changes are probably connected with the fact that an argon MIP never covers the complete cross-section of the discharge tube in any kind of cavity (fig. 3). Some of the aerosol will therefore miss the plasma, or at least the hot part of it. This could imply that the degree of atomization depends on the nature and the concentration of the matrix. In addition we must also take account of changes in the excitation conditions, since the matrix effects for atomic lines may differ appreciably from those for the ionic lines of the corresponding elements.

For assessing the usefulness of an argon MIP as an excitation source for simultaneous multi-element analyses of solutions we make a comparison with other plasma-excitation sources: the ICP, mentioned above, and the capacitively coupled microwave plasma (CMP). To give an ideal comparison the spectrometric conditions and the rate at which these samples can be injected should be identical. As an approximation we use the results obtained from a comparative investigation [27] on an argon ICP and a nitrogen CMP, in which the spectrometric conditions and the injection rates were not very different from those used in our investigation on an argon MIP.

Table VII shows a comparison of the minimum detectable concentrations for a number of emission lines obtained with each of the three excitation sources. Now for the *atomic* emission lines, the minimum detectable concentrations for the three excitation sources are of the same order of magnitude. An important feature of the ICP, however, is that highly sensitive *ionic* emission lines are available for many elements. For these elements, therefore, minimum detectable concentrations one to three orders of magnitude better can be obtained. There is no such advantage with MIP, or with CMP [27].

The relative standard deviation of the net line intensity (the observed gross intensity less that of the background) is usually found to be 2 to 3% at concentrations that are about a hundred times greater than the minimum detectable concentration. This value is better than that for the CMP (3 to 5%), but not so good as that for the ICP (about 1%). The MIP also occupies an intermediate position for the matrix effects. These are smaller than with the CMP but larger than with the ICP. Apart from this, the ICP has the highest matrix tolerance and the greatest linear range.

[23] C. A. Bache and D. J. Lisk, Anal. Chem. 39, 786, 1967.
[24] W. R. McLean, D. L. Stanton and G. E. Penketh, Analyst 98, 432, 1973.
[25] D. N. Hingle, G. F. Kirkbright and R. M. Bailey, Talanta 16, 1223, 1969.
K. Fallgatter, V. Svoboda and J. D. Winefordner, Appl. Spectrosc. 25, 347, 1971.
H. Kawaguchi, M. Hasegawa and A. Mizuike, Spectrochim. Acta 27B, 205, 1972.
[26] P. W. J. M. Boumans and F. J. de Boer, Spectrochim. Acta 32B, 365, 1977.
[27] P. W. J. M. Boumans, F. J. de Boer, F. J. Dahmen, H. Hoelzel and A. Meier, Spectrochim. Acta 30B, 449, 1975.

Combination with an electrothermal atomizer

Since a stable helium plasma can be generated at atmospheric pressure with the new cavity, an MIP can be combined with an electrothermal atomizer. For the substrate of the evaporator we used a 2 mm thick 'cord' of pyrolytic graphite. This cord is stretched between two graphite support electrodes, mounted in a Pyrex glass chamber; see *fig. 9*. The helium gas flows into the atomization chamber through the support electrodes and then to the discharge tube. The sample for analysis (1 to 3 µl) is applied to the cord by a syringe. *Fig. 10* shows the complete experimental arrangement. *Fig. 11* shows an electrothermal atomizer in operation.

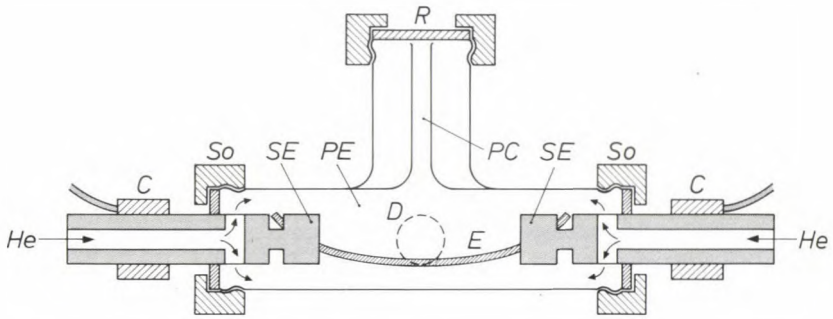


Fig. 9. Diagram of the cross-section of an electrothermal atomizer. *E* evaporation substrate consisting of a 'cord' of pyrolytic graphite. *SE* support electrodes of graphite. *R* thin rubber seal. *PC* Pyrex guide channel for syringe needle. *PE* Pyrex evaporation chamber. *So* Sovirel screw couplings. *D* discharge tube perpendicular to plane of diagram. *C* connection to electrical supply.

A cord of pyrolytic graphite is particularly suitable for use in the atomizer. It can rapidly be raised to a temperature of 2000 to 3000 °C, and requires relatively little power (about 300 W). It has a fairly long life, typically 500 atomization cycles or more. The temperature is easily regulated by means of the applied voltage, and measurements with an optical pyrometer have shown that there is a linear relation between the temperature and the voltage. Injecting a very small quantity of liquid (1 µl) on to a given substrate with a syringe is usually a rather inaccurate proces (20 to 30% error). However, when the graphite cord has been heated to about 80 °C, it is no longer water-repellant and can in fact be wetted. The injected droplet is 'soaked up' by the cord; this gives a much more reproducible sample injection, to within 2 to 3%.

In the well-known application of a graphite tube as an electrothermal atomizer in atomic absorption spectrometry, the evaporation of the sample, the atomization of compounds and the atomic absorption all occur at the same place. In the combination of an electrothermal atomizer and an MIP the atomization and excitation occur at different places. Since the atomization and the excitation can also take place in the plasma, the temperature of the graphite cord is less critical and need not be so high. This is illustrated in *fig. 12*, where the emission intensity of manganese and lead is plotted as a function of the temperature of the cord. If the temperature is high enough for complete evaporation, the emission intensity is independent of the temperature.

The construction of the electrothermal atomizer described here, with the gas introduced through the support electrodes, is designed as far as possible to avoid 'dead' spaces. As a result the analysis signal as a function of time has the shape of a symmetrical peak, as shown in *fig. 13* for a signal originating from 10 ng of manganese. This means that both the height and the area of the signal response can be used as a measure of the concentration of the element to be determined in the injected solution. In *fig. 14* the peak height is plotted on a log-log scale as a function of the

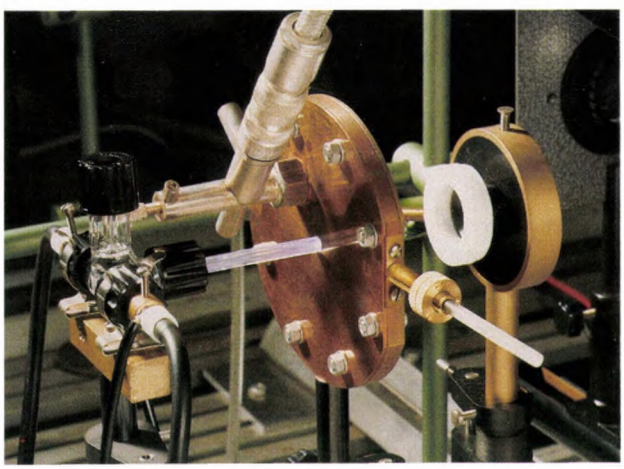


Fig. 10. Photograph of a helium MIP with the cavity at the centre, the electrothermal atomizer on the left and a lens with protective glass on the right.

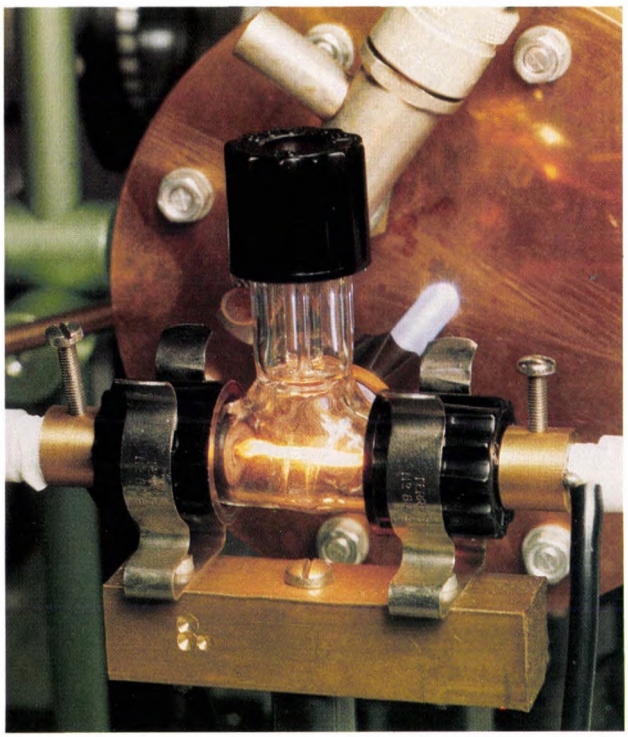


Fig. 11. An electrothermal atomizer with incandescent graphite cord.

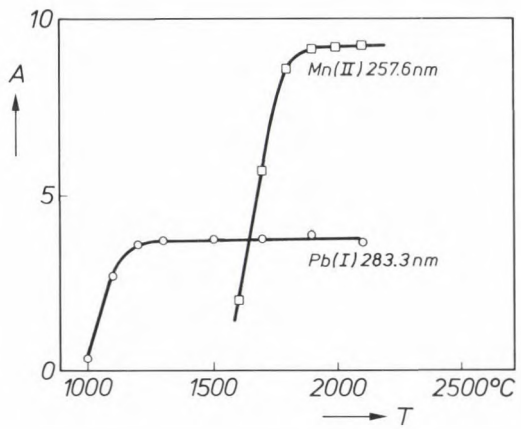


Fig. 12. Peak area A (arbitrary units) of emission signals of manganese and lead in an MIP, as a function of the incandescence temperature T of the graphite cord in an electrothermal atomizer. The increase of A with temperature is the result of increasing atomization. If the temperature is high enough for complete atomization, A does not continue to increase with temperature.

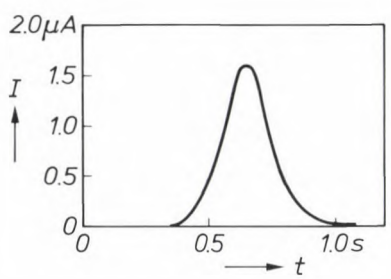


Fig. 13. Peak current I of the emission signal at 257.6 nm originating from 10 ng of manganese, as a function of time t after switching on the electrothermal atomizer.

injected quantity for three elements. It can be shown from the slopes of the curves that the relation between peak height and injected quantity is almost linear. *Table VIII* gives the detection limits that have been determined with these calibration lines for a number of elements.

As in the combination with a nebulizer, matrix effects can also occur here. *Fig. 15* shows how a matrix

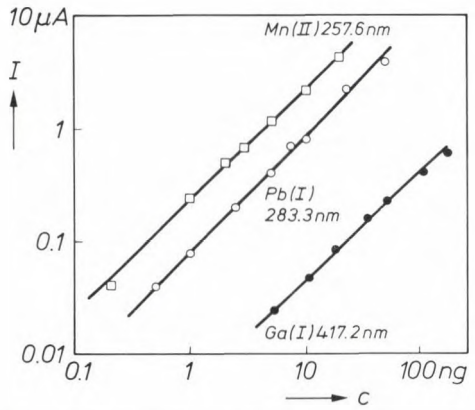


Fig. 14. Peak current I of emission signals of manganese, lead and gallium, as a function of the quantity c injected in 1 μl of solution.

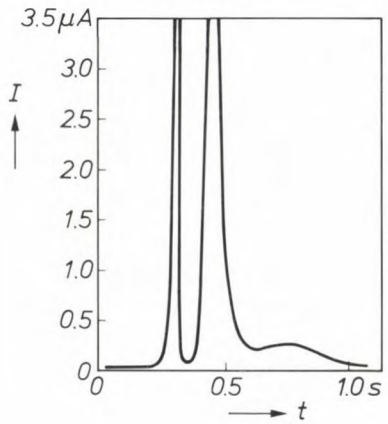


Fig. 15. Peak current I of the emission signal at 257.6 nm originating from 10 ng of manganese in a matrix with 1% of KCl, as a function of time t after switching on the atomizer. Because of the presence of KCl the peak shape differs considerably from the 'ideal' shape in fig. 13.

effect can distort the shape of the peak. In this figure the measuring signal again originates from 10 ng of manganese, but now in a solution that also contain 1% of potassium chloride. With such a distorted signal the peak height can no longer be used for determining the concentration. The distortion of the signal here is connected with the difference in evaporation and atomization between manganese and potassium chloride. When the graphite cord is heated up, the large surplus of potassium chloride present evaporates at a much lower temperature than the manganese. In this process

Table VIII. Wavelength λ of the analysis line and detection limit c_{min} for a number of elements, determined with a combination of an MIP and an electrothermal atomizer.

Element	λ (nm)	c_{min} (ng)
Ga(I)	417.2	0.01
Ca(I)	422.7	0.3
Ca(II)	393.4	0.02
Mn(I)	403.1	0.01
Mn(II)	257.6	0.001
Pb(I)	283.3	0.007
S(II)	564.0	10
Cl(II)	479.5	1
Br(II)	478.6	1

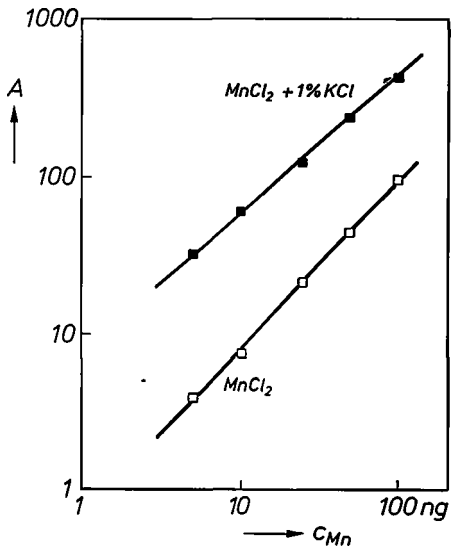


Fig. 16. Peak area A (arbitrary units) of the emission signal of manganese at 257.6 nm, as a function of the injected quantity c_{Mn} for pure $MnCl_2$ solutions and for $MnCl_2$ solutions with 1% of KCl as matrix. In spite of the marked distortion of the peak shape (fig. 15) due to the presence of KCl, there is a clear relation between the manganese content and the peak area.

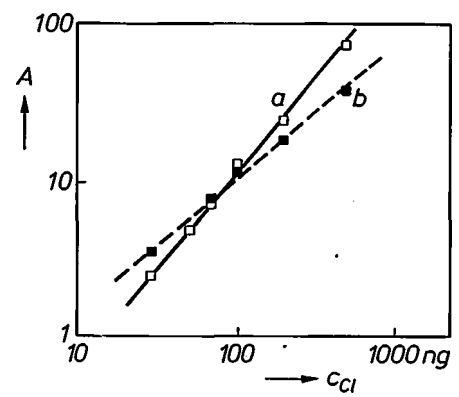


Fig. 17. Peak area A (arbitrary units) of the emission signal of chlorine at 479.5 nm, as a function of the quantity of chlorine c_{Cl} injected in 1 µl of solution, for sodium chloride (*a*) and *ortho*-chloroacetanilide (*b*). Here again the matrix has a marked influence on the strength of the signal.

some of the manganese is probably carried along by the rapidly evaporating matrix, causing the sharp irregular peak at the start of the signal. Nevertheless, there remains a unique relation between the integrated signal and concentration; see *fig. 16*. It can be seen here that the net signal has been considerably enhanced by the presence of potassium chloride.

We were able to recover the ideal peak shape in fig. 13 by heating the graphite cord in two stages: the matrix is first cautiously evaporated at a temperature of 900 °C, and once this evaporation has been completed the graphite cord is heated to over 2000 °C to obtain the emission signal of manganese. In this way matrix effects can largely be suppressed by means of suitable temperature programming of the graphite cord. This procedure has to be re-optimized separately for every matrix, however, as in atomic absorption spectrometry.

The combination of a helium MIP and an electrothermal atomizer can also be used for elements such as sulphur and the halogens. Although the detection limits (measured in ng) for these elements are not so good as those for most metals (Table VIII), they compare favourably with the values obtained for halogens with other analytical methods (measured in µg). In the determination of these elements it is again necessary to take matrix effects into account. An example is shown in *fig. 17*, in which the integrated signal for two chlorine compounds, sodium chloride and *ortho*-chloroacetanilide, is plotted on a log-log scale as a function of the injected quantity of chlorine. The slopes obtained differ considerably and indicate an appreciable deviation from the linear relation.

Status of the MIP-AES method in analytical spectrometry

The value of the MIP as an excitation source for AES analyses is determined not only by the analytical capabilities but also by such factors as the size of the equipment, cost, the nature of the samples for analysis and the minimum quantity required [28].

Since a stable helium plasma can be generated at atmospheric pressure with the new cavity, the MIP-AES method shows great promise for sequential or simultaneous multi-element analysis of gas mixtures, e.g. after separation with a gas chromatograph. Organic and organometallic compounds can be detected rapidly with appreciable sensitivity and selectivity; the equipment is relatively simple, takes little power and only a small amount of plasma gas is required.

[28] P. W. J. M. Boumans, Philips tech. Rev. 34, 305, 1974.

If it is necessary to perform multi-element analyses of solutions in the macro range (50 µl to a few ml), the MIP, like the ICP, has to work with argon. Even when the new cavity is used, the MIP may often compare unfavourably with the ICP as an excitation source. This is because the coupling to the nebulizer presents more problems, the minimum detectable concentrations are generally worse, the matrix effects are more significant, and the system does not readily tolerate high matrix concentrations (e.g. 10 mg/ml). These disadvantages may more than outweigh the advantages of lower power, lower gas quantities and lower cost.

The situation is different when it comes to analysing micro-quantities of solutions (1-10 µl). Here the helium MIP with electrothermal evaporation has more to offer. In particular, elements such as sulphur and the halogens are readily detected; this is virtually impossible with the ICP. If it is necessary to determine micro-quantities of metals, the MIP has to compete with flameless atomic absorption spectrometry; this is more sensitive, but is not suitable for simultaneous multi-element analysis. As in atomic absorption spectrometry, the analyses are sometimes upset by matrix effects, which make appropriate temperature programming necessary. In principle, this may have to be different for every element to be detected in a particular matrix, which is something of a limitation to the multi-element nature of the system.

In conclusion, we can say that the MIP-AES method occupies a position of its own in analytical spectrometry and is a welcome addition to other analytical methods.

Summary. A microwave-induced plasma (MIP) can be used in atomic emission spectrometry as an excitation source for sequential or simultaneous multi-element analysis. The plasma gas is helium or argon. With the resonant cavities commonly used earlier, the useful application of the MIP was limited because it was not possible to generate a stable plasma in helium and wet argon at atmospheric pressure. This can be done with the new cavity described here, since there is a much better transfer of microwave energy to the plasma. The excitation mechanisms that play an important part in an MIP can be deduced from the wavelengths of the observed emission lines and the associated excitation energies, and from the energy-level diagrams and ionization energies of the different elements. This explains why a helium MIP can also be used for elements with emission lines of high excitation energy. These elements cannot be determined from line emission spectra with an argon MIP. A helium MIP at atmospheric pressure can be used for simple, rapid and sensitive analysis of gas mixtures, e.g. after separation with a gas chromatograph. The MIP is less promising for the analysis of solutions when it has to be combined with a nebulizer and an argon plasma must be used. On the other hand, for analysing micro-quantities of solutions (1-10 µl) successful use can be made of a helium plasma at atmospheric pressure by combining it with an electrothermal atomizer. This combination is particularly useful when elements such as sulphur and the halogens have to be determined.

Calculation and design of electro-mechanical devices

B. Aldefeld

In designing electro-mechanical devices it is desirable to have a quantitative knowledge of the physical processes that determine their behaviour. These include the distribution of the magnetic field, the effects of eddy currents and the changes in the magnetic circuit caused by the motion of the armature. Computer calculations can be used to calculate the effects of a large number of parameters before a device is actually constructed. At the Philips Hamburg laboratories a computer program has been developed that can be used to provide a rapid analysis of a large number of variations to a design. The program provides accurate predictions of the dynamic behaviour of these devices.

Introduction

Electro-mechanical devices (sometimes called electromagnetic devices) are driven by a magnetic field generated by an electric current flowing in a coil combined with a suitably shaped iron circuit. The magnetic field exerts forces on an armature of magnetizable material, so that mechanical energy is produced. Examples of such devices are electro-mechanically actuated valves, brakes, clutches and printers. The calculation and design of such devices is the subject of this article.

When electro-mechanical devices have to be newly designed or improved, preliminary calculations are necessary to show whether the expected performance will meet the target specifications. Accurate calculations can reduce considerably the number of experimental investigations to be made. Such experiments are not only expensive, but are extremely time-consuming. They cannot be omitted altogether, of course, for there are always some problems that defy calculation. Questions concerning matters such as durability, resistance to wear and material processing are not amenable to mathematical treatment, and experimental tests on prototypes are therefore indispensable.

As an example of an electro-mechanical device, *fig. 1* shows an electromagnet used in a dot-mosaic printer. A printer of this type consists of a number of needles that can be separately actuated so as to produce a character on paper. The electromagnets are required to meet the following specifications: an available mechanical energy of about 1 mJ, a high speed of action (repetition rate between 1000 and 1500 Hz), accurate timing, high electro-mechanical efficiency, and a long life. And of course the manufacturing costs should be as low as possible.

The conventional methods of computing the performance of such devices are based on simplified analytical expressions. These have their uses for obtaining a general insight into idealized cases and for estimating the variation of certain quantities. However, even for the simple geometry of the magnet in fig. 1, these classical methods do not provide sufficiently accurate solutions. This is due in the first place to the nonlinearity of the magnetization curve of iron. This means that the permeability is dependent on the induction (magnetic flux density) B, so that the contribution of the iron to the total magnetic field varies in a manner that is not easily predictable. Another problem is that there are eddy currents if the iron is not laminated. Their intensity varies with space and time, and depends on several other factors such as the shape of the iron circuit, the electrical conductivity and the magnetic permeability.

These problems can be tackled by means of suitable numerical methods, known as finite-element and finite-difference methods, which are also widely applied in other fields, for instance in the calculation of mechanical stresses [1]. These methods are based directly on the fundamental differential equations describing the problem, making it possible to obtain results with an accuracy that depends only on the computer time available.

Dr B. Aldefeld is with Philips GmbH Forschungslaboratorium Hamburg, Hamburg, West Germany.

For problems of a dynamic nature, however, which we are concerned with here, even the capabilities of modern computers are inadequate when it comes to dealing with arbitrary three-dimensional problems. The computer time required would be far too long, for reasons that will become obvious when the numerical method used is described. Nevertheless, many of the problems encountered in practice — such as geometries with rotational symmetry — can in fact be solved because they are mathematically two-dimensional.

The investigation and development of numerical methods of calculation for magnets with rotational symmetry has been the subject of several years of work at Philips Forschungslaboratorium Hamburg. The objective was the development of a computer program that could be used as a design aid. This program, called EDDY (EDdy currents, DYnamic behaviour) was completed towards the end of 1978. The methods on which it is based, the problems it can solve and some results will be described in this article.

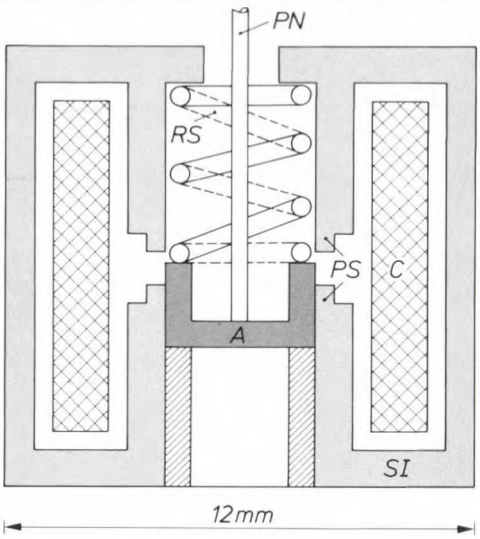


Fig. 1. Electromagnetic actuator used in a dot-mosaic printer. *C* actuating coil. *SI* soft-iron core. *PS* pole pieces. *A* armature with print needle *PN*. *RS* return spring, which returns the armature to its initial position after the needle has touched the paper. The magnet is energized by a current or voltage pulse lasting about 0.5 ms. The printed characters consist of patterns of single dots.

Basic equations

The behaviour of an electromagnet that is excited by a variable current is largely described by two differential equations, one relating to the magnetic field, the other to the motion of the armature. Other relations also enter into the picture. If the magnet is excited from a voltage generator, the current dependent self-inductance of the system has to be calculated. If the magnet is excited by a very high current, heat generation may be a limiting factor. This is often the case in practice.

Magnetic field

To compute the magnetic field throughout the entire magnetic circuit we make use of the vector potential, $\boldsymbol{A}$, which can be derived at every point from the current distribution [2]. There is some analogy between this vector potential, generated by currents, and the electrostatic potential, which arises from charges.

The derivation of an electric field-strength from the electrostatic potential is well known. The potential is defined at a point x,y,z as

$$\phi(x,y,z) = \frac{1}{4\pi\varepsilon_0} \int \frac{\varrho(V)}{r_V(x,y,z)} \mathrm{d}V,$$

where ε_0 is the permittivity of free space, $\varrho(V)$ is the charge density in the volume element $\mathrm{d}V$ and r_V is the distance between $\mathrm{d}V$ and the point x,y,z. Similarly the vector potential $\boldsymbol{A}(x,y,z)$ is defined as

$$\boldsymbol{A}(x,y,z) = \frac{\mu_0}{4\pi} \int \frac{\boldsymbol{j}(V)}{r_V(x,y,z)} \mathrm{d}V,$$

where μ_0 is the permeability of free space and $\boldsymbol{j}(V)$ is the current density in the volume element $\mathrm{d}V$. From the electrostatic potential the electric field-strength $\boldsymbol{E}$ can be derived by differentiation with respect to the coordinates of position, or by applying the gradient operator:

$$\boldsymbol{E} = -\operatorname{grad}\phi = -\frac{\partial\phi}{\partial x}\boldsymbol{n}_x - \frac{\partial\phi}{\partial y}\boldsymbol{n}_y - \frac{\partial\phi}{\partial z}\boldsymbol{n}_z.$$

The magnetic flux density $\boldsymbol{B}$ can be derived in much the same way from $\boldsymbol{A}$ by means of the relation $\boldsymbol{B} = \operatorname{curl}\boldsymbol{A}$. The analogy between ϕ and $\boldsymbol{A}$ ends here, because ϕ is a scalar quantity and $\boldsymbol{A}$ is a vector.

Provided the current distributions are not too complicated, $\boldsymbol{A}$ has the same direction everywhere as the current. In the magnets considered in this article we are only concerned with currents in the tangential direction of the cylindrical coordinates and therefore $\boldsymbol{A}$ will also be tangential, that is to say the only non-zero component of $\boldsymbol{A}$ is A_φ.

In practical situations it is not easy to calculate the vector potential from the definition given above. However, we know from Maxwell's equations [3] that

$$\operatorname{curl}\boldsymbol{B}/\mu = \boldsymbol{j},$$

[1] J. H. R. M. Elst and D. K. Wielenga, The finite-element method and the ASKA program, applied in stress calculations for television picture tubes, Philips tech. Rev. **37**, 56-71, 1977.
[2] See for example J. Volger, Vortices, Philips tech. Rev. **32**, 247-258, 1971.
[3] R. L. Stoll, The analysis of eddy currents, chapter 1; Clarendon Press, Oxford 1974.

and hence

$$\operatorname{curl}\left(\frac{1}{\mu}\operatorname{curl} A\right) = j,$$

where μ is the magnetic permeability of the material at the point where we wish to calculate A, and j is the current density. If this is converted into cylindrical coordinates a differential equation is obtained, from which A can be calculated. As stated earlier, the only non-zero component of A is the tangential A_φ. The differential equation to be solved finally becomes [3]

$$\frac{\partial}{\partial r}\left(\frac{1}{r\mu}\frac{\partial}{\partial r}rA_\varphi\right) + \frac{\partial}{\partial z}\left(\frac{1}{r\mu}\frac{\partial}{\partial z}rA_\varphi\right) = \\ = -j_\varphi + \sigma\left(\frac{\partial A_\varphi}{\partial t} + v_z\frac{\partial A_\varphi}{\partial z}\right). \quad (1)$$

The last two terms on the right-hand side of the equation have been added to describe the effect of the eddy currents and the movement of the armature (σ is the electrical conductivity of the magnetic material and v_z is the velocity of the armature). Note that the differential equation (1) is nonlinear, since the permeability μ depends on the flux density, which in turn is a function of the vector potential.

The differential equation describes the magnetic field completely, because once it has been solved for the vector potential, all other quantities of the electromagnetic field can be found from simple relations. For instance, the components of the magnetic flux density B are given by

$$B_r = -\frac{1}{r}\frac{\partial}{\partial z}rA_\varphi,$$

and

$$B_z = \frac{1}{r}\frac{\partial}{\partial r}rA_\varphi.$$

These relations follow directly from Maxwell's equations.

Motion of the armature

If we assume that the armature can only move in the z-direction, its motion is given by the equation

$$\frac{\partial^2 z}{\partial t^2} = \frac{1}{m}(F_m + F_s + F_f), \quad (2)$$

where m is the mass of the armature; F_m, F_s and F_f denote the axial components of the magnetic force, the spring force and the frictional force, in that order.

The force exerted on the movable parts of the magnetic circuit by the magnetic field can be evaluated from the flux-density distribution. This force is given in integral form by the surface integral [4]:

$$F_m = \int\left\{\frac{1}{\mu_0}B(B\cdot n) - \frac{1}{2\mu_0}B^2 n\right\}\mathrm{d}S, \quad (3)$$

where n is the vector normal to the surface element $\mathrm{d}S$. (The surface of integration can be freely chosen, provided that it completely encloses the movable part and does not intersect other magnetizable parts.)

Method of solution

General procedure

The equations given above are solved by time-stepping methods. Time-stepping means that solutions to the equations are calculated at small time intervals, with each new solution based on the previous one. Since the differential equations (1) and (2) are interdependent, they have to be solved simultaneously. (The equations are connected ('coupled') via the magnetic force and the position and velocity of the armature.)

If only static fields and forces are to be computed, equation (2) is not relevant, and the right-hand side of equation (1) only contains the current density, which is constant. In the time-dependent case the solution procedure starts at the time $t = 0$, when the vector potential is zero throughout the geometry and the armature is stationary in its initial position. Then, using small time steps, the two differential equations are solved alternately until the final state is reached.

To improve the accuracy of the solution, a prediction-correction method can be used at each time step. In the prediction phase the position of the armature is predicted by solving eq. (2) on the assumption that the magnetic force has remained the same as the computed value of the previous step. The vector potential for this new position is then calculated from eq. (1), and from this the new magnetic force is calculated from eq. (3). In the correction phase the differential equation for the position of the armature is solved again, now using the *new* value for the magnetic force. The solution of the ordinary differential equation (2) can now easily be obtained by means of well-known, fast and reliable methods, such as the Runge-Kutta procedure [5]. Such standard methods are not available for solving the nonlinear partial differential equation (1). Here the search for improved techniques is still in progress, with particular emphasis on the reduction of the computer time required. The methods used in the EDDY program are described in the next section.

Finite-difference method

The essential feature of the numerical methods for the solution of a partial differential equation is that the equation is discretized, which means that solutions are only calculated for discrete points in space and time. The discretized analogue of the differential equation is a set of algebraic equations, which is amenable to solution by computer. Once the solution for the discrete points has been obtained, an interpolation procedure can be used to obtain the solution for the region between the points. The density of the points determines the accuracy to which the differential equation is approximated. A denser distribution gives an improved accuracy, but also requires more computer time because the calculation has to be made for more points. These methods are known as finite-difference methods.

Discretization

To obtain the analogues of the partial derivatives with respect to space, the geometry is considered to be subdivided into a number of elements by means of a rectangular mesh. The nodes of the mesh define the points at which the value of the vector potential will be calculated. In the case of rotational symmetry, it is sufficient to discretize only half of the cross-sectional area.

As an example, *fig. 2* shows such a mesh for the electromagnet of fig. 1, together with the outline of the iron parts and the coil.

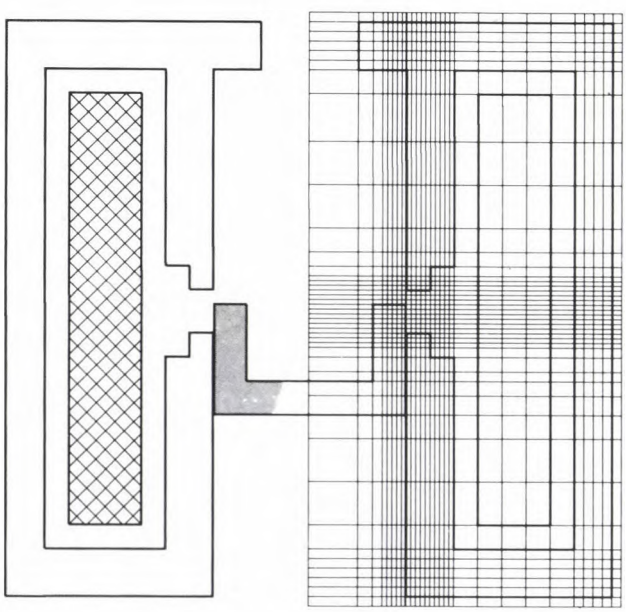


Fig. 2. Example of a finite-difference mesh. The spacings of the mesh (the mesh lengths) are variable, so that smaller spacings can be used in regions where a more accurate solution is required, as in the air-gap region. All elements are assigned values for the permeability, conductivity, velocity, current density and magnetic flux density.

A mesh of this type consists of horizontal and vertical lines that have points of intersection over the whole region. It can be set up very easily, since only two sets of lines have to be specified, and a fast algorithm can be found for modifying it in accordance with the time-varying position of the armature. In addition, the resulting equations have a simple structure, which also simplifies the numerical solution. A disadvantage of this simple mesh is that shapes with curved surfaces — other than those with rotational symmetry — are difficult to approximate.

A detailed description of the mathematical procedure for obtaining the discretized form of the differential equation would be beyond the scope of this article, but a procedure for a simplified case can be explained briefly. Let us consider the differential equation

$$\frac{\partial^2 A}{\partial x^2} + \frac{\partial^2 A}{\partial y^2} = 0, \tag{4}$$

which is a highly simplified form of eq. (1) in Cartesian coordinates, with the right-hand side set equal to zero. Let the mesh in this simple case have a constant mesh length h, as shown in *fig. 3*. The approximation to the first derivatives of the vector potential can then easily be obtained from the ratios of the differences. For the spatial derivative at a point half-way between nodes $l+1$, m and l,m, for example, this gives:

$$\frac{\Delta A}{\Delta x} = \frac{1}{h}(A_{l+1,m} - A_{l,m}).$$

From these first derivatives an approximation to the second derivatives at the point l,m can be obtained in the same way, giving for the second derivative to the x-coordinate:

$$\frac{\Delta^2 A}{\Delta x^2} = \frac{1}{h^2}(A_{l+1,m} + A_{l-1,m} - 2A_{l,m}),$$

and a corresponding relation for $\Delta^2 A/\Delta y^2$. The finite-difference analogue of the differential equation (4) is then

$$\frac{1}{h^2}(A_{l+1,m} + A_{l-1,m} + A_{l,m+1} + A_{l,m-1} - 4A_{l,m}) = 0, \tag{5}$$

or

$$A_{l,m} = \tfrac{1}{4}(A_{l+1,m} + A_{l-1,m} + A_{l,m+1} + A_{l,m-1}).$$

An expression such as (5) is called a five-point finite-

[4] B. Aldefeld, Forces in electromagnetic devices, Comm. Proc. COMPUMAG Grenoble 1978, paper 8.1.

[5] R. Zurmühl, Praktische Mathematik für Ingenieure und Physiker, 5th edition, chapter 27; Springer, Berlin 1965.

difference equation, since five nodes of the mesh are used in the expressions.

In the case of the more complicated differential equation (1), a different procedure for the space discretization is more suitable because the vector potential cannot be differentiated at the boundaries of the material [6]. In addition, the time domain also has to be discretized by introducing discrete time instants at intervals Δt. The result is again a five-point finite-difference equation which, somewhat simplified, has the form:

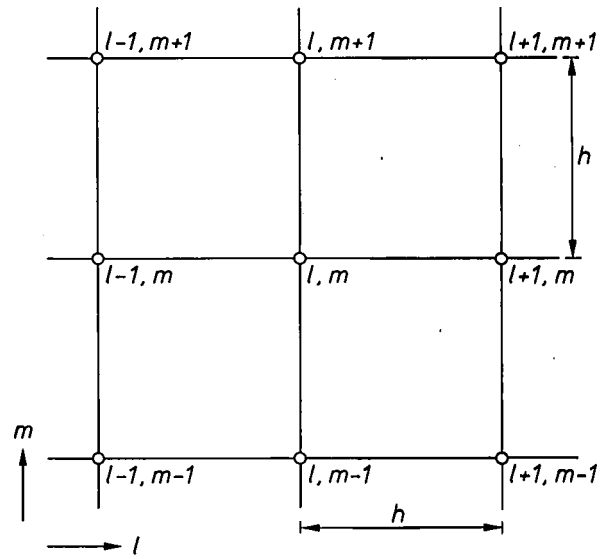


Fig. 3. Section of a finite-difference mesh with a constant mesh length h. The indices l,m are used to identify the nodes.

$$A_{l,m}^{t+\Delta t} = aA_{l,m}^{t} + i_{l,m}^{t+\Delta t} + bA_{l+1,m}^{t+\Delta t} + cA_{l-1,m}^{t+\Delta t} + dA_{l,m+1}^{t+\Delta t} + eA_{l,m-1}^{t+\Delta t}. \quad (6)$$

Here t denotes the time for which a solution has already been found, and $t + \Delta t$ denotes the new time for which the solution is not yet known. The term $i_{l,m}$ is the discretized current; a, b, c, d and e are coefficients that depend on the local values of permeability, conductivity, velocity and mesh length. Since there is a five-point equation for each node of the mesh, there are as many equations as nodes. The equations are all interdependent and can be assembled into a matrix equation.

In practical applications the number of equations can be considerable. Consider, for instance, a problem in which the region of interest is subdivided into 30 elements per coordinate direction: for rotational symmetry there are then 900 equations with 900 unknowns to be solved at each time step. In an arbitrary three-dimensional geometry, this becomes 27 000 equations with 27 000 unknowns at each time step. An estimate of the computer time that would then be needed shows that this method of solution would be too cumbersome for most practical applications.

Solving the finite-difference equations

For solving large numbers of equations with many unknowns fast methods are available, such as Gaussian elimination. These methods are widely used in engineering calculations for buildings. However, methods of this type cannot be applied here, because the coefficients of the finite-difference equation (6) are not constant but depend on the local flux density. Instead, a solution is obtained by means of an iterative process in which both the vector potential and the coefficients are calculated. The method used in the EDDY program is one of the class known as relaxation methods. In these the vector potential is calculated from the five-point equation (6), and the entire mesh is scanned repeatedly until eq. (6) is satisfied to a certain accuracy for all nodes. The algorithm consists in the following steps:

1. Calculate the coefficients of eq. (6) from the values of the flux density of the preceding time step.
2. Apply eq. (6) successively to all nodes of the mesh to obtain the values of the vector potential $A_{l,m}$.
3. Calculate the flux density from the vector potential for all cells of the mesh. Calculate the local permeability from the flux density and the magnetization curve and determine the coefficients.
4. Go back to step 2 and continue.
5. Stop the procedure as soon as the differences between successive values of the vector potential are below a specified limit.

In the EDDY program a number of improvements that help to accelerate the convergence of the iterative process have been introduced. First, instead of applying the five-point equation to each single node, all the five-point equations for all nodes on one mesh line are treated simultaneously in a single matrix equation. This equation is solved by using a fast elimination algorithm [7], thus giving new values for all the nodes on a line simultaneously. The iteration consists in repeated application of this procedure to all the lines of the mesh. This kind of technique is called 'line iteration'.

Secondly, a method called 'overrelaxation' is used. This method is based on the fact that in each iteration the values obtained for the vector potential, $A_{l,m}$, approach the solution only slowly by a relatively small amount $\Delta A_{l,m}$. The increment $\Delta A_{l,m}$ can therefore be multiplied by a factor ω greater than 1, so that the solution is approached more rapidly. (It can be shown that the factor ω, the 'overrelaxation' factor, must

always be less than 2, otherwise the process would not converge.)

Thirdly, use is made of Maxwell's first equation, which states, in its integral form, that the line integral of the magnetic field-strength B/μ around a closed path must be equal to the number of ampere-turns enclosed by the integration path. Accelerated convergence is achieved by evaluating the line integral after each iteration and correcting all values of the vector potential in such a way that the integral criterion is satisfied.

It should be noted that in the method described it is necessary to take some precautions to ensure the convergence of the iterative process. Trouble may arise from the extreme nonlinearity of the magnetization curves of ferromagnetic materials. This is because relatively small differences in the flux density between successive iterations can cause large differences in the permeability and hence in the coefficients of eq. (6). This may give rise to oscillations or even to divergence. If this effect is observed, it can be corrected by keeping the changes in the coefficients below a certain limit, e.g. by introducing an 'underrelaxation factor' ($\omega < 1$), by analogy with the overrelaxation factor.

The EDDY computer program

The EDDY program was developed from the numerical methods outlined above, and is applicable to systems with rotational symmetry. The solution provides flux-density and eddy-current distributions, lines of force, energy losses, magnetic forces, motion of the armature and static temperature distributions. The program is written in standard FORTRAN and consists of 70 subroutines with a total of about 10 000 lines of FORTRAN text. The maximum number of nodes it can handle is 5000, and the core-storage requirement is 250 kbytes, with overlay techniques.

The user normally has to specify the initial input data, which is summarized in *Table I*. If static magnetic fields and forces only are to be calculated, some of this data is irrelevant and need not be specified. If the static temperature distribution of the system is to be calculated, the thermal conductivities must be specified instead of the magnetization curves. The computer time required for finding a solution varies considerably with the problem to be solved, and depends on several factors such as the shape of the magnetization curves and the number of nodes and time steps. For the problems solved so far, the time required on a Philips P1400 has varied from five minutes to two hours for the calculation of the dynamic behaviour.

Application

The EDDY program has mainly been applied in the design of electromagnets used in dot-mosaic printers. Some of the many calculations performed will be described below to illustrate the capacities of the program.

Static fields and forces

As a first step in design calculations for an electromagnet, it is advisable to investigate static fields and forces alone. In this way a general picture of several features can be obtained with a minimum of computer time, and the information thus obtained can be used for estimating some of the dynamic quantities if the eddy-current effect is not too pronounced. Some examples of such calculations are considered in figs 4, 5 and 6. *Figs 4a* and *4b* show the lines of force and the flux-density distribution for the electromagnet of fig. 1. Together these figures provide a complete picture of the magnetic field, as they display both its intensity and its direction. It should be noted that fig. 4*b* gives a relatively imprecise representation and that more accurate flux-density values must be obtained from the tabular print-out.

Table I. Input data required for the EDDY computer program.

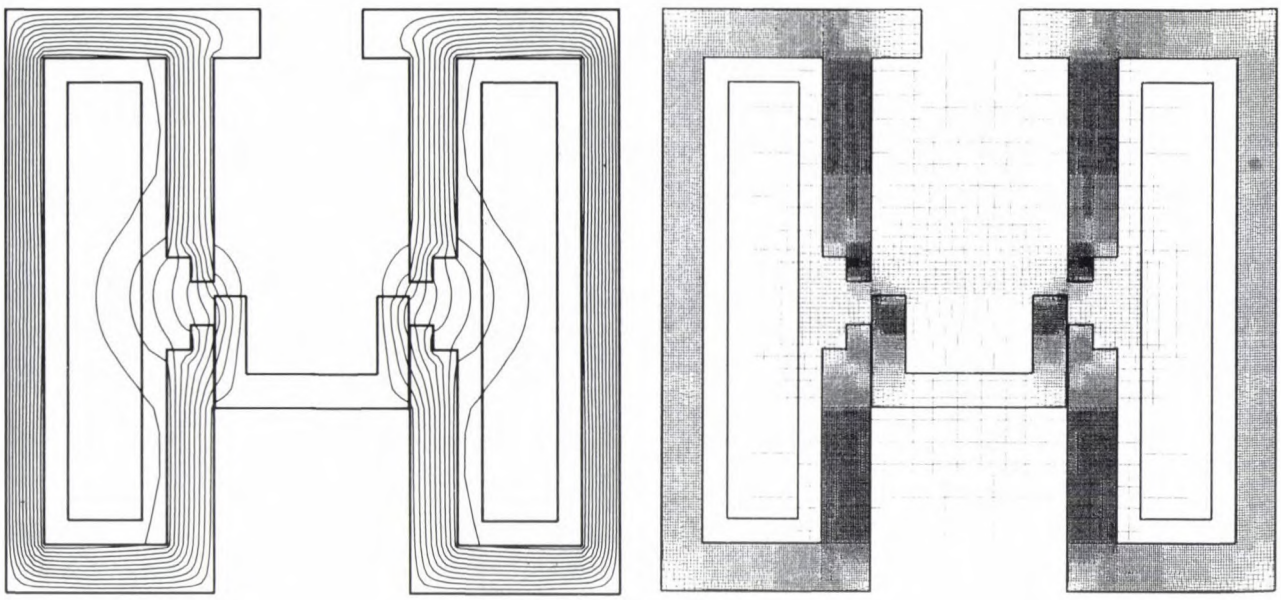
Configuration	dimensions and shape of magnetic circuit and coil; number of turns and electrical resistance of coil; frictional force, initial spring tension and spring constant; electrical conductivity; magnetization curves
Excitation	shape of current pulse at coil terminals
Discretization	finite-difference mesh and time steps
Parameters	relaxation factors; convergence criteria; max. permissible number of iterations [*]
Output	specifications relating to results of calculations (drawings, tables)

[*] These parameters need only be specified if they differ from the values given in the program.

Fig. 5 gives an enlarged view of the air-gap region of fig. 4*a* and shows the pattern of the lines of force in this part of the magnet more clearly.

The variation of the magnetic force as a function of the displacement of the armature is shown in *fig. 6* for two values of the excitation current. These curves are

[6] R. L. Stoll, The analysis of eddy currents, chapter 8; Clarendon Press, Oxford 1974.
[7] R. S. Varga, Matrix iterative analysis, chapter 6.4; Prentice-Hall, Englewood Cliffs 1962.



a *b*

Fig. 4. Lines of force (*a*) and flux-density distribution (*b*) in the electromagnet of fig. 1 for constant-current excitation. In this example the current density in the coil is 10^8 A/m^2. The iron is magnetically saturated in the central region, where the flux density B is well above the saturation value (2.15 T).

useful because they show where the armature should be placed to achieve maximum force and hence maximum acceleration.

Dynamic behaviour; field-penetration rate

The calculation of the dynamic behaviour provides the designer with nearly all the information he needs about the electromagnetic and mechanical characteristics of the device. The dynamic behaviour of high-speed electro-mechanical devices is largely determined by the rate at which the magnetic field penetrates into the ferromagnetic material. This rate determines the maximum frequency at which the device can be operated.

To understand the details of this penetration process, it should first be studied for a simple case. A suitable geometry for such an example is that of semi-infinite space. The penetration process in this case is quasi-one-dimensional, because the field penetrates into the material in the form of a plane wave. A one-dimensional study of this kind can be performed by applying the EDDY program to a simple cylindrical configuration in which the penetration depth is small compared with the other dimensions.

A result of such a study is shown in *fig. 7* for the case of field excitation by a step pulse. As expected, the curves show that the penetration process depends closely on the properties of the material. In high-speed devices it is particularly important that the rate of flux penetration into the material should be very high. In this respect, material *2* is to be preferred, as

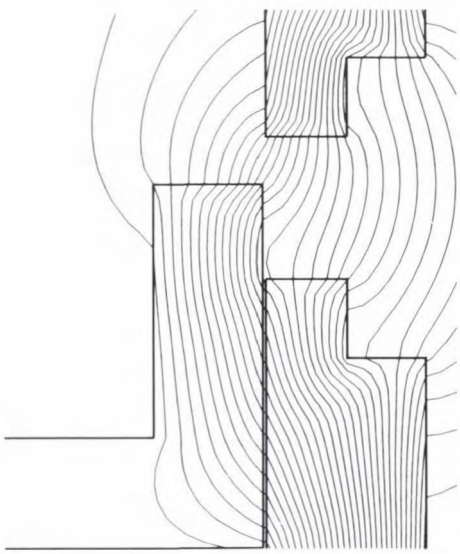


Fig. 5. Enlarged view of the air-gap region of fig. 4*a*.

the rate of penetration is about three times higher than in material *1* for the same external field-strength (or the same current density in the coil). However, the maximum flux density is lower with material *2*, and this feature favours material *1*. The best choice in a particular case depends on the other specifications the design has to meet [8].

[8] B. Aldefeld, Electromagnetic field diffusion in ferromagnetic materials, Proc. IEE **125**, 278-282, 1978.

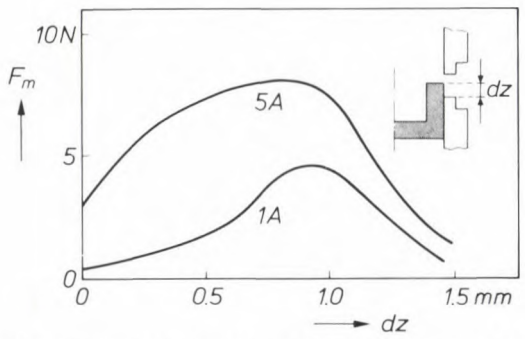


Fig. 6. Static force F_m as a function of the armature displacement dz of the electromagnet in fig. 1. The excitation currents of 1 A and 5 A correspond to current densities in the coil of 2×10^7 and 10^8 A/m². The force has its maximum value when the front of the armature is near the corner of the upper pole piece. The relatively small increase in the force as the current increases from 1 A to 5 A is due to saturation of the magnetic circuit, particularly for the larger displacements of the armature.

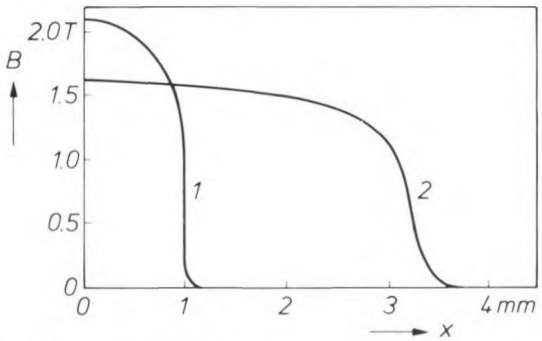


Fig. 7. Penetration of the magnetic field into a ferromagnetic material that fills a semi-infinite space, for the case where the field at the surface of the material is suddenly switched on and then remains constant. (In this example $H = 4.8\times10^4$ A/m). The curves — flux density B as a function of depth x — which relate to different materials, represent the flux waves at $t = 0.1$ ms after the field has been switched on. Material *1* has a saturation flux density of 2.15 T and an electrical conductivity of $6\times10^6\ \Omega^{-1}\ m^{-1}$. For material *2* the corresponding values are 1.7 T and $7\times10^5\ \Omega^{-1}\ m^{-1}$. The rate at which the field penetrates into the material is determined by the eddy currents produced.

Examples

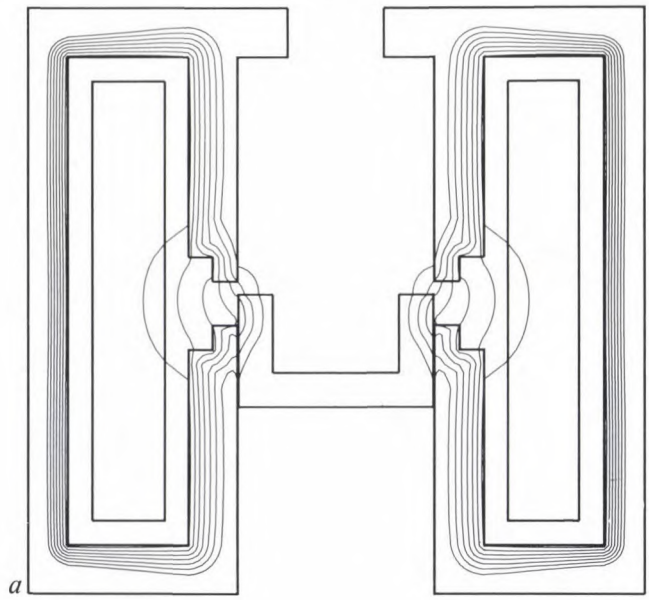
To optimize a design a single calculation is not sufficient, of course. In general, several calculations with varied input data have to be carried out. As an example, some results of a dynamic calculation are shown in figs 8, 9 and 10. The results all relate to the configuration in fig. 1. Further details are given in *Table II*.

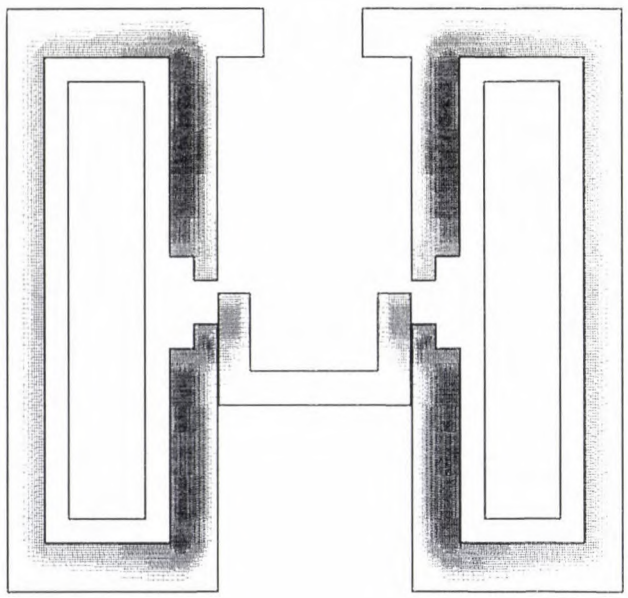
One of the most remarkable characteristics of the dynamic process is that the electromagnetic field penetrates into the ferromagnetic material in the form of a wave with a steep front (*fig. 8*), and that the time required for the penetration is comparable with the duration of the excitation pulse. This behaviour comes about because the effects of eddy currents are by no means negligible, since the eddy currents are of the same order of magnitude as the currents in the

Table II. Data for calculating the dynamic behaviour of the magnet in fig. 1.

Saturation flux density	2.15 T
Electrical conductivity	$6\times10^6\ \Omega^{-1}\ m^{-1}$
Maximum displacement of the armature (up to return point)	0.5 mm
Energy loss after impact	90% of impact energy
Spring constant	800 N/m
Initial spring tension	0.25 N
Frictional force	0.1 N
Number of turns of coil	300
Electrical resistance of the coil	4 Ω
Mass of armature	300 mg

Fig. 8. Lines of force (*a*) and eddy-current density (*b*) in the iron core and the armature of the dot-mosaic printer, 0.18 ms after the current has been switched on.





coil. Another characteristic is that eddy currents still continue to flow after the current in the coil has dropped to zero. The eddy currents produce a field pattern in which most of the lines of force do not link with the coil but remain inside the iron (*fig. 9*). The energy for these currents is taken from the energy that was previously stored in the magnetic field.

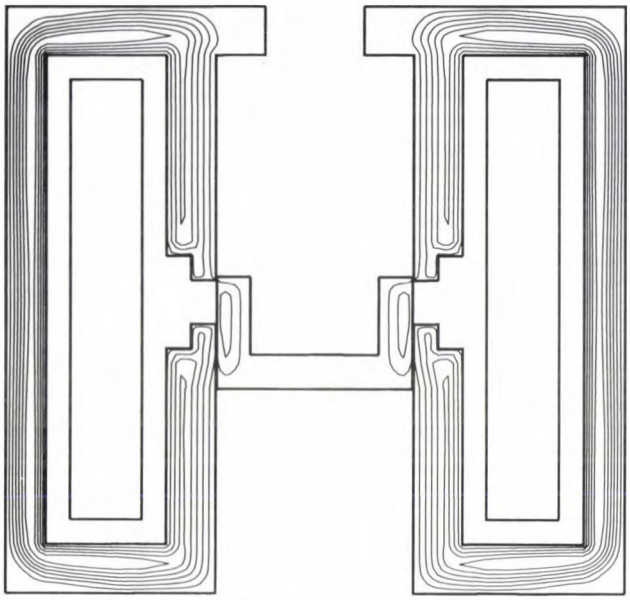


Fig. 9. Lines of force 0.75 ms after switching on. The coil current has already dropped to zero, so that the eddy currents alone determine the field pattern.

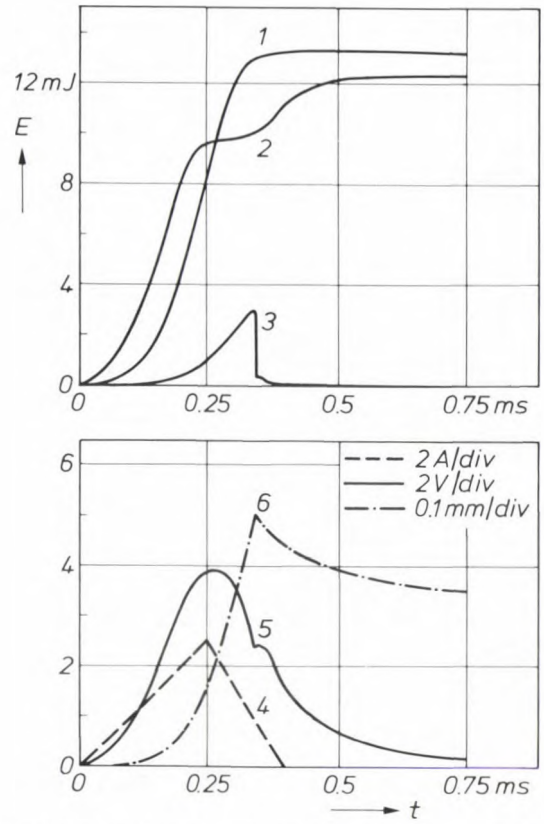


Fig. 10. Some of the curves describing the dynamic behaviour of the dot-mosaic printer during the first 0.75 ms after the current is switched on. *1* Energy dissipated in coil. *2* Energy dissipated by eddy currents in the magnetic material. *3* Mechanical energy of the armature. *4* Current in the coil. *5* Magnetic force. *6* Displacement of the armature.

The importance of the eddy-current effect may be seen from a comparison of the curves in *fig. 10*. The power dissipated in the coil and the eddy-current losses are of nearly equal magnitude, but the mechanical energy is about four times lower. A reduction in the eddy current will therefore be very effective in improving the electro-mechanical efficiency.

Another undesired effect of eddy currents is that they cause the magnetic force to decay relatively slowly after the excitation. This decelerates the armature during its return phase, which in turn decreases the maximum possible repetition rate at which the printer can operate. One remedy would be to use a stronger return spring. This would at the same time have the positive effect of decreasing the mechanical energy, which is too high in the present example; about 1 mJ is sufficient to actuate a print needle.

The example given makes it clear that a large number of parameters can be varied. The EDDY program can be used for calculating large numbers of differently combined parameters, enabling an optimum design to be found within a short time. This design can then be constructed and tested.

The work described here was sponsored by the German Federal Ministry for Research and Technology under contract HH-PHI/103. Responsibility for the contents of this article rests with the author.

Summary. The behaviour of electro-mechanical devices can be calculated to a high accuracy by means of numerical computer methods. The EDDY program developed at Philips Forschungslaboratorium Hamburg is based on a finite-difference method. The nonlinear partial differential equation that describes the problem is solved by means of a relaxation technique with accelerated convergence. EDDY can be used for calculating the static and dynamic behaviour of devices that have rotational symmetry. The solution provides the distribution of the magnetic field and the eddy currents, the energy consumption, the losses, the magnetic forces and the motion of the armature. EDDY has mainly been used in the design of dot-mosaic printers.

The electrodeposition of aluminium

J. F. M. van de Berg, T. E. G. Daenen, G. Krijl and R. E. van de Leest

Until recently the electrodeposition of aluminium on a conducting substrate was only very rarely put to practical use. This was mainly because the deposition had to be carried out in a highly flammable organic solvent. Electrochemical studies at Philips Research Laboratories have now led to the discovery of a much less dangerous solvent. The laboratories are now cooperating with the Centre for Metal Chemistry and Coatings — part of the Philips Plastics and Metalware Factories (PMF) — in an investigation of the industrial feasibility of the process in a pilot production scheme, and are trying out various likely applications for electrodeposited aluminium.

Introduction

Aluminium is one of the most abundant metals on the surface of the Earth. It is widely distributed and only occurs naturally in the combined state. The metal can be obtained by electrolysis of an aluminium compound. A condition here is that there should be no water in the electrolysis solution, since because of the non-noble nature of aluminium it is not the trivalent aluminium ions that give up their charge but the hydrogen ions (protons) of the water. The electrolysis process that has been widely used since 1886 for producing aluminium is based on molten cryolite (Na_3AlF_6) in which aluminium oxide (Al_2O_3) has been dissolved. This process still accounts for about 95% of the total production of aluminium. It is not suitable for coating other metals with a layer of aluminium, since the electrolysis has to be performed at 1000 °C, a temperature at which aluminium metal is in the liquid state.

Aluminium can be electrodeposited, however, if the electrolysis is carried out in a solution of an aluminium compound in an *organic* solvent. The advantage of this is that the electrolysis can take place at room temperature. Good results have been obtained, for example, with a solution of aluminium chloride ($AlCl_3$) and lithium hydride (LiH) or lithium aluminium hydride ($LiAlH_4$) in diethyl ether [1]. However, industrial application is difficult because diethyl ether is extremely flammable and volatile.

In the investigation described in this article we carried out experiments with solutions of $AlCl_3$ and $LiAlH_4$ in the cyclic ether tetrahydrofuran (THF), which can also be used for the electrodeposition of aluminium. Tetrahydrofuran is less volatile than diethyl ether and has a higher flash-point. However, there is still some risk in large-scale operations with tetrahydrofuran. We have used electrochemical methods to enable us to study the reaction mechanisms that govern the electrodeposition of aluminium. With the knowledge gained in this way we were able to find a solvent that considerably reduced the fire risk. This in turn led to an increased interest in the application of electrodeposited aluminium.

In this article we shall first discuss the electrodeposition of aluminium, paying some attention to the reaction mechanism of the deposition and the choice of organic solvent. Then we shall say something about the properties of electrodeposited aluminium and discuss several possible applications.

Ing. J. F. M. van de Berg, G. Krijl and Dr R. E. van de Leest are with the Philips Plastics and Metalware Factories, Eindhoven; Dr T. E. G. Daenen is with Philips Research Laboratories, Eindhoven.

[1] D. E. Couch and A. Brenner, J. Electrochem. Soc. 99, 234, 1952.

The electrodeposition of aluminium

Fig. 1 is a diagram of a simple arrangement that can be used for the electrodeposition of aluminium. A special system of airlocks prevents air from coming into contact with the electrolyte. Above the electrolyte there is an inert atmosphere of argon, which is heavier than air and therefore acts as a buffer between the electrolyte and the air. The anode is made from engineering-grade aluminium. To prevent the incorporation of small particles of metal or impurities liberated in the anodic dissolution of this grade of aluminium, the anodes are protected by polypropylene screens during the electrolysis.

A solution of 1.2 mol/l of $AlCl_3$ and 0.4 mol/l of $LiAlH_4$ in THF has a resistivity of about 300 Ωcm. This is a low value for an organic electrolyte. Nevertheless, so much heat is generated at high current densities that the solution has to be cooled.

In electrolysis the rate of growth of the aluminium layer depends on the current density. At 10 mA/cm^2 the growth rate for aluminium is 12.4 μm per hour. The cathodic current efficiency is 100%. At a current density of 100 mA/cm^2 an aluminium layer with a maximum thickness of 10 μm can be deposited in an unstirred solution. At a higher current density or with a thicker layer the growth becomes irregular and dendrites may be formed.

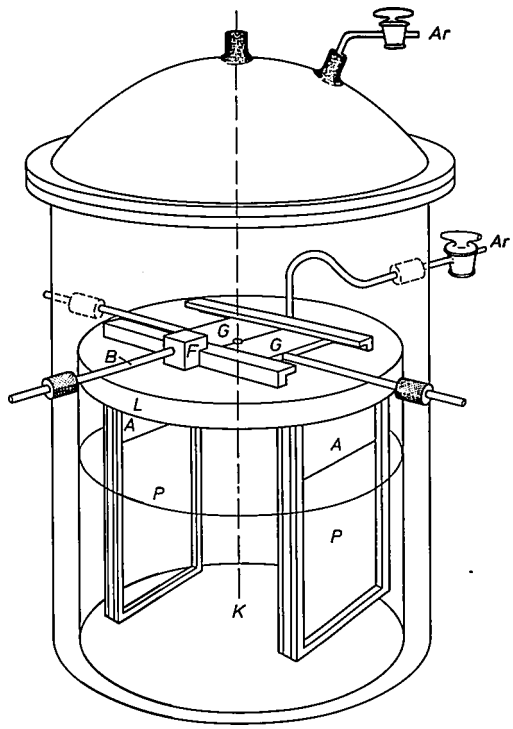


Fig. 1. Diagram of a simple arrangement for the electrodeposition of aluminium. The electrolysis takes place in a glass vessel that is isolated from the surrounding air by an outer vessel filled with argon. The electrolysis vessel is sealed by a Teflon cap *L* in which there is a feed-through *F* for the anode connection and a tube for the supply of argon. The opening at the centre of *L* is for the introduction of the electrolyte (3 litres) and the cathode *K*, indicated by a dashed line. This opening can be closed with the glass plates *G*, which can be operated from outside. The anodes *A* are attached to the underside of *L*; they are protected by polypropylene screens *P*. *B* copper rod for the anode connection.

With the arrangement shown in fig. 1 we have been able to deposit aluminium layers corresponding to 45 ampere-hour equivalents and of very high quality on various materials. This was done during a period of longer than four months, without having to change the solution. The equipment could be used for electroplating substrates with a maximum size of 9×9×6 cm. Larger substrates can now be electroplated in an improved arrangement corresponding to the block diagram of *fig. 2*. The equipment was mainly constructed from glass and Teflon (polytetrafluoroethylene, PTFE). The improvements are related to cooling and changing the electrolyte, filtering out impurities and the process control.

Reaction mechanism

From electrochemical measurements we have been able to show that two compounds play an important part in the electrodeposition of aluminium in the solution described above [2]. The first is aluminium hydride, which is formed in the reaction between aluminium chloride and lithium aluminium hydride:

$$AlCl_3 + 3LiAlH_4 \rightarrow 4AlH_3 + 3LiCl.$$

The second is the complex ion $AlCl_2.(THF)_4^+$, which is formed when aluminium chloride dissolves in tetrahydrofuran:

$$2AlCl_3 + 4THF \rightarrow AlCl_2.(THF)_4^+ + AlCl_4^-.$$

Aluminium hydride is the compound that is reduced to aluminium metal at the electrode. The $AlCl_2.(THF)_4^+$ ions react with the liberated H^- ions and thus ensure the formation of fresh aluminium hydride:

$$\begin{array}{ccc} 3e + AlH_3 & \rightarrow & 3H^- + Al \\ \uparrow & & \downarrow \\ 4THF + 2Cl^- + AlH_3 & \leftarrow & 3H^- + AlCl_2.(THF)_4^+ \end{array}$$

Since the regeneration reaction proceeds very rapidly, the AlH_3 concentration in the neighbourhood of the electrode remains constant in this cyclic process. The $AlCl_2.(THF)_4^+$ ions can easily provide the mass transport, since they are present in a high concentration in the solution. These ions are not reduced to aluminium metal; this would require a much higher voltage than is necessary for the reduction of AlH_3: electrolysis of a solution of $AlCl_3$ alone in tetrahydrofuran does not yield aluminium.

The rate of deposition is considerably affected by the occurrence of a cyclic reaction mechanism of this type. This can be deduced from measurements of the electric current as a function of the applied voltage, for example. *Fig. 3* shows current/voltage curves for a solution of aluminium hydride in tetrahydrofuran, both with and without aluminium chloride. In the solution with no $AlCl_3$ there is no possibility of regenerating AlH_3 via the cyclic mechanism. The rate of deposition is then determined by the rate of mass transport, as can be seen from the presence of the diffusion plateau in the current/voltage curve. AlH_3 is however regenerated in the solution with $AlCl_3$, and there is no diffusion plateau in the measured curve.

The occurrence of the cyclic reaction mechanism has a beneficial effect on the quality of the deposited layer of aluminium. This is illustrated in the scanning-electron-microscope photographs in *fig. 4* of the layers obtained with the two solutions. In the solution with no aluminium chloride a concentration gradient has arisen as a result of the electrode reaction, and consequently the nuclei that protrude into the depleting solution will grow more quickly than the others. The irregularities of the electrode surface are therefore magnified. The deposited aluminium layer has a rough and spongy appearance and dendrites will appear if growth continues. The layer deposited from the solution containing aluminium chloride looks quite different. Because of the cyclic reaction mechanism the aluminium-hydride concentration at the surface remains constant and the aluminium layer produced is compact, in spite of the higher current density.

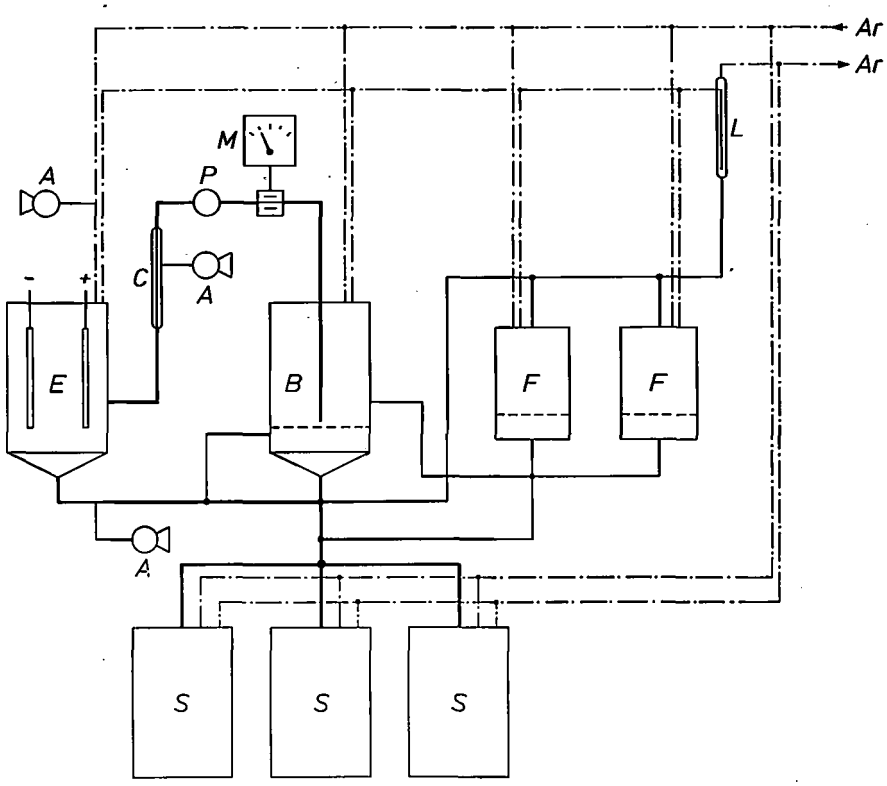


Fig. 2. Block diagram of an improved arrangement for the electrodeposition of aluminium. The order in which the substrate for electrodeposition and the electrolyte are introduced is the reverse of that in the arrangement of fig. 1: here the substrate is put into the empty electrolysis vessel *E* first, and then the electrolyte is drawn into *E* from a storage vessel *S* or the buffer vessel *B*. The connecting tubes for the electrolyte and also for the supply and removal of argon are shown schematically. *F* filtration vessels for removing deposits formed in the electrolyte during the electrolysis. *M* meter indicating the electrical conductance. *C* cooling. *L* cold trap. *A* alarm circuit. *P* pump.

Solvent

Proceeding from the knowledge we had gained of the reaction mechanism, we investigated several organic solvents to assess their usefulness in the electrodeposition of aluminium. It was found that the same reaction mechanism also operates in other ethers. However, most ethers are not suitable for use in an industrial electrolysis process, since they are too flammable. This is especially so for diethyl ether, as we saw in the introduction, and less so for tetrahydrofuran.

[2] T. E. G. Daenen, Nature 280, 378, 1979.

A much more suitable solvent is diethylene glycol dimethyl ether ('diglyme'), which is much less volatile and less flammable; see *Table I*. In earlier investigations diglyme was found unsuitable, since aluminium chloride only dissolves in it sparingly [3]. However, in our investigations of the reaction mechanism we found that the important factor was not so much the solubility of aluminium chloride as the solubility of the combination of aluminium chloride and aluminium hydride. The solubility of this combination in diglyme was found to be sufficiently high, and the first results with this solvent were very promising.

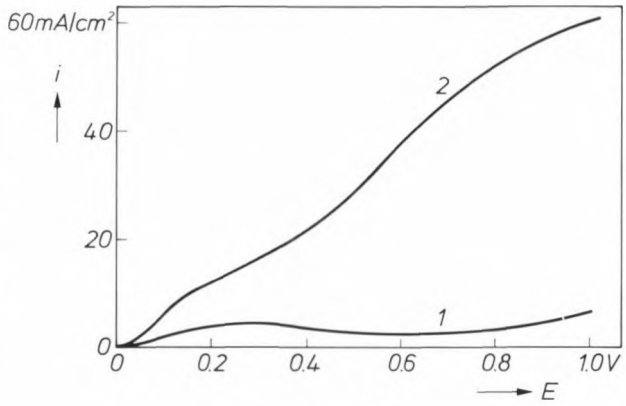


Fig. 3. Current density i as a function of the applied voltage E for the electrodeposition of aluminium at room temperature. Curve *1* refers to a solution of 0.17 mol/l of a complex of AlH_3 and triethylamine (in the ratio of 1 : 2) in tetrahydrofuran (THF) containing $LiAlH_4$. Curve *2* refers to the same solution, but now with added $AlCl_3$.2THF (at 0.44 mol/l). In this case there is no diffusion plateau arising from the occurrence of a cyclic reaction mechanism due to the presence of $AlCl_3$.

Properties and possible applications of electrodeposited aluminium

The possible applications of electrodeposited aluminium are related to the properties of aluminium metal and its oxide. The metal is relatively light, is a good electrical conductor, is non-magnetic and is not toxic. It reflects visible and ultraviolet radiation better than most other metals. Only gold and silver give a better reflection in the infrared region. Aluminium is almost unaffected by air, water and most acids. This is because a thin transparent oxide film is formed immediately the metal is exposed to air. This oxide film protects the aluminium underneath it and slowly increases in thickness to a maximum value between 3 nm and 10 nm.

Thicker oxide layers can be obtained by anodizing. When electroplated aluminium is anodized, a hard wear-resistant oxide layer is formed, which is resistant to most chemicals. This layer can be used as a very good electrical insulator, as a simple coating that can easily be dyed, and for corrosion protection.

We have electrodeposited aluminium on various kinds of steel, brass, nickel, nickel-iron, aluminium, phosphor-bronze, molybdenum, gold and carbon. Any desired layer thickness between 1 and 300 μm can be obtained by selecting appropriate values for the current density and the electrolysis time. Electrodeposited aluminium is ductile and its hardness lies in the range from 30 to 80 HV (Vickers), so that it can readily be finished mechanically, e.g. by polishing, or even milled or turned with a diamond-tipped tool.

In cooperation with J. M. Oomen of the Mechanical Development Group at Philips Research Laboratories we have made mirrors of very high optical quality by machining electrodeposited aluminium. *Fig. 5* compares the quality of a surface of electrodeposited aluminium turned with a diamond-tipped tool with the quality of a surface of engineering-grade aluminium machined in the same way. This application requires a very good adhesion to the substrate.

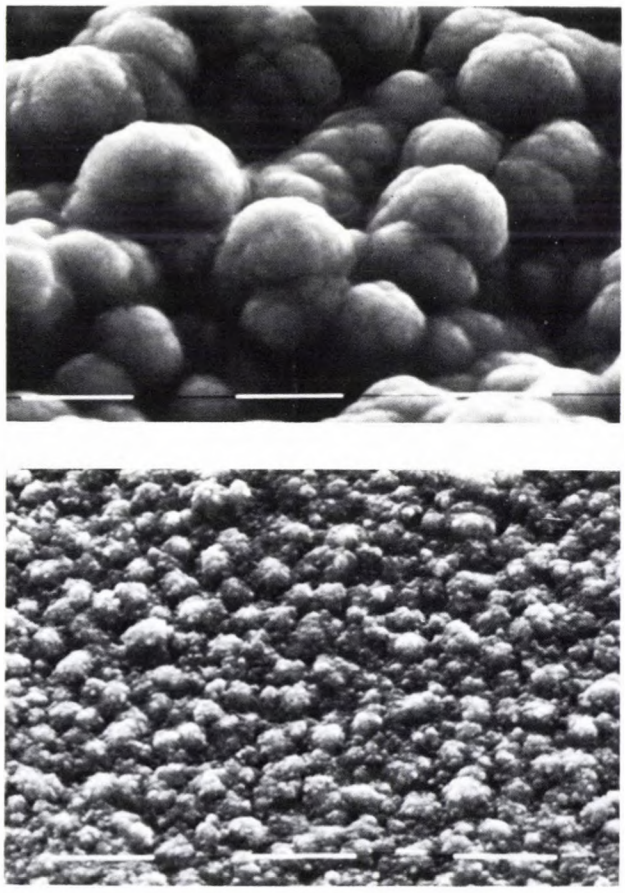


Fig. 4. Scanning-electron-microscope photographs of aluminium layers of comparable thickness, deposited from the solutions of fig. 3 at room temperature. The length of a black or white strip is 10 μm. *Above:* from the solution with no $AlCl_3$, at a current density of 3.5 mA/cm² for 90 min. The layer has a rough and spongy surface. *Below:* from the solution with $AlCl_3$, at a current density of 21 mA/cm² for 15 min. This layer is compact in spite of the higher current density.

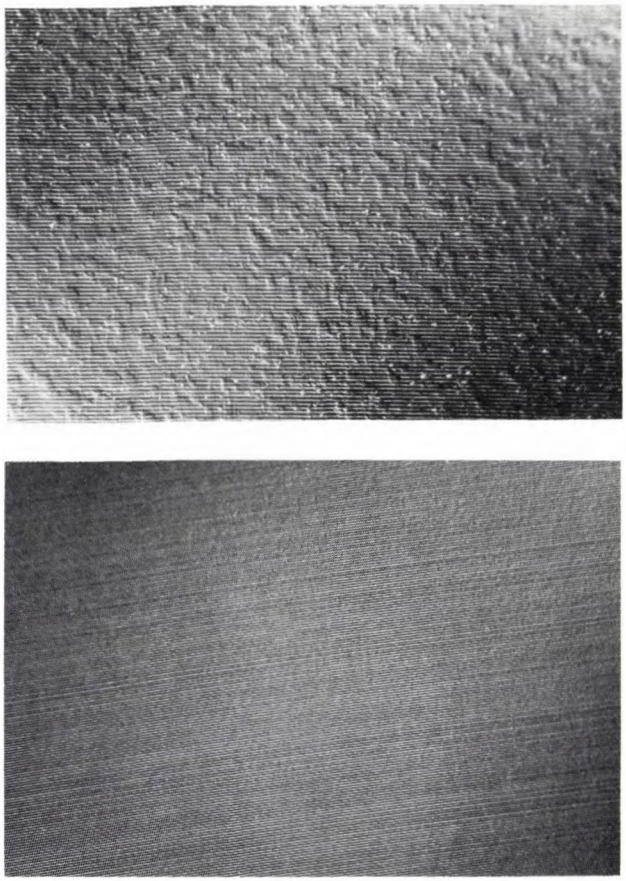


Fig. 5. Photographs of aluminium surfaces that have been turned with a diamond-tipped tool (magnification 35×). *Above:* surface of engineering-grade aluminium. *Below:* surface of 35 μm electrodeposited aluminium on engineering-grade aluminium. The quality of this surface is clearly much better.

[3] J. H. Connor and A. Brenner, J. Electrochem. Soc. **103**, 657, 1956.

[4] 'Macor' is a kind of ceramic glass that can be machined easily.

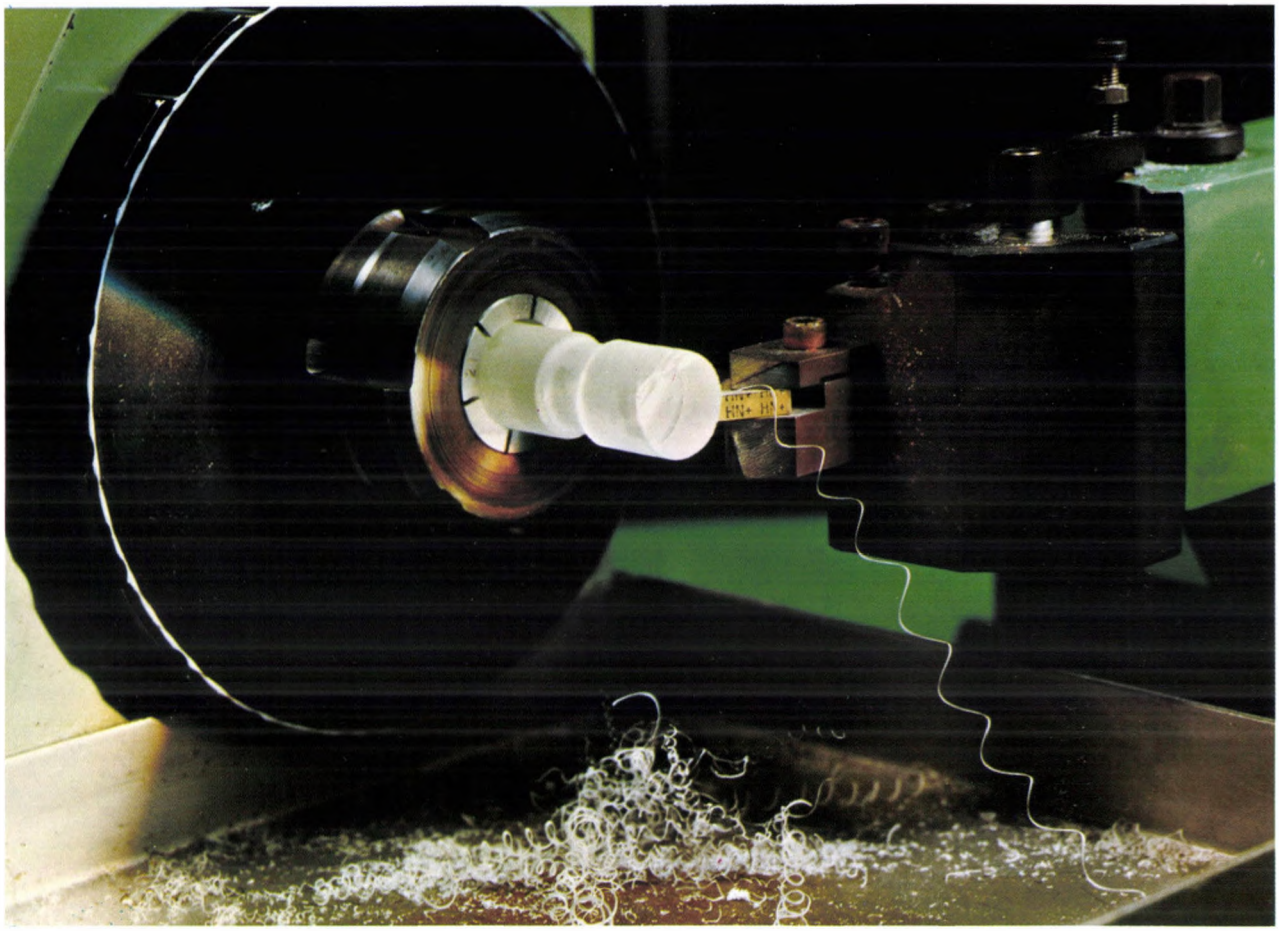
Table I. Boiling-point and flash-point of three organic solvents that can be used in the electrodeposition of aluminium.

Solvent	Boiling-point (°C)	Flash-point (°C)
Diethyl ether	35	−40
Tetrahydrofuran	66	−17.5
Diethylene glycol dimethyl ether	162	70

We achieved this by using an intermediate layer of copper and tin. A reflecting aluminium coating can also be applied to *non*-conducting substrates such as glass or 'Macor' [4], provided an intermediate conducting layer has been applied first.

Now that aluminium can be electrodeposited in an acceptably safe electrolytic process to form a layer with many useful properties, all kinds of applications can be envisaged. We have to remember, of course, that the method of electrolysis is more demanding than conventional electrodeposition. Before we can start to think of mass production, for example for protecting steel against corrosion, it will have to be possible to operate the process very rapidly. At the moment, it looks as though the process is particularly suitable for coating high-value products that can benefit from the specific properties of aluminium and aluminium oxide.

Summary. In the electrodeposition of an aluminium layer from a solution of $AlCl_3$ and $LiAlH_4$ in an organic solvent such as diethyl ether or tetrahydrofuran a cyclic reaction mechanism arises that has beneficial effects on the properties of the layer. Until recently the electrodeposition of aluminium has been associated with a fire hazard, which has severely limited industrial applications. The use of diethylene glycol dimethyl ether ('diglyme') appreciably reduces this fire hazard.



Transparent single-point turning of glass

Glass technologists have for years been aware of a need for making glass objects of optical quality by using cutting techniques like the long-established ones used for other materials. Machining on a lathe, for example, would be a rapid way of producing aspherical glass surfaces. One advantage of aspherical lenses in optical systems is that they require far fewer elements than comparable completely spherical systems.

Turning glass surfaces of optical quality on a lathe sets some exacting requirements on the cutting process and on the accuracy of operation of the lathe. We were fortunate in that we were able to make use of an optical precision lathe developed at Philips Research Laboratories [1]. This lathe has hydrostatic bearings and a speed-controlled hydraulic motor. The first experiments showed that glass surfaces with a good optical finish cannot be obtained at room temperature: the resulting surface finish of the various types of glass was rough and matt; the cutting tools showed excessive wear and the waste material came away as a powder.

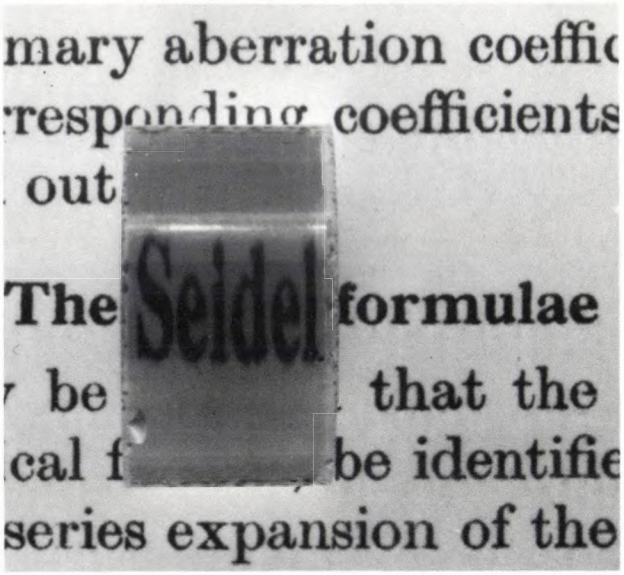
The title photograph shows a glass workpiece being turned on a lathe, with a continuous 'shaving' being removed. To achieve such a good cutting action we heat the glass, while turning it, to a temperature close to the 'American' softening point, giving it a viscosity of about $10^{6.6}$ Nsm^{-2}. This temperature varies from 700 to 800 °C for the optical glasses that we have successfully machined in this way.

At such high temperatures diamond tends to turn into graphite and also has a high thermal conductivity, so that the tool tip can become a cold point in the machining process. We therefore turned our

[1] H. J. J. Kraakman and J. G. C. de Gast, Philips tech. Rev. **30**, 117, 1969.

attention to other materials. The best results were obtained with a cutting tool consisting of a thin hafnium-nitride layer on a hard metal substrate.

To obtain a transparent surface for the glass it is necessary to control its temperature relatively accurately, to within a range of about 20 °C. (The conventional machine settings such as speed and rate of feed also have to be correctly chosen, of course.) The workpiece is heated by two burners. One is a slit-burner, which heats the workpiece to a few tens of degrees below the required machining temperature. The second is a point-burner; this burner is placed close to the tool and its gas feed is controlled automatically in relation to the measured temperature to provide appropriate local heating. The adjacent photograph shows an example of a transparent cylinder lens of BaLF 5 glass produced by this method.



R. Brehm
K. van Dun
J. Haisma
J. C. G. Teunissen

Ing. R. Brehm and Dr Ir J. Haisma are with Philips Research Laboratories, Eindhoven; K. van Dun and Ing. J. C. G. Teunissen were formerly with Philips Research Laboratories; K. van Dun is now with the Philips Personnel and Industrial Relations Department, Eindhoven, and Ing. J. C. G. Teunissen is now with the Philips Audio Division, Eindhoven.

Thin-film strain-gauge transducers

K. Bethe and D. Schön

A strain-gauge transducer converts a force into a mechanical strain that is detected by way of a change in electrical resistance. Several types of such transducers have been developed since the invention of the strain gauge. This article deals with transducers in which the strain gauges are made by the deposition of thin films. Advances in thin-film technology, as achieved for example in the integration of electronic circuits, have considerably improved the capabilities and performance characteristics of strain-gauge transducers.

Introduction

For measuring physical quantities such as mass, force, torque, acceleration, flow and pressure it is very convenient to use transducers that give an electrical signal. Such transducers contain an elastic element that is deformed by an applied force, and an electrical sensor that detects the deformation. The electrical signal can then be amplified and transmitted over long distances. It can be fed into a control system or processed by a computer and stored, and it can be displayed in an easily readable form. Electrical transducers can be small, and continue to be reliable even after long periods of use.

A sensor can convert a mechanical deformation into an electrical signal by a variation in inductance or capacitance, or by a change in electrical resistance as in the well-known strain gauge [1]. Although strain gauges are not as sensitive as inductive and capacitive sensors [2], they have so many technical and economic advantages (*Table I*) that their applications continue to increase. To determine a strain-induced change in electrical resistance, four strain gauges are connected to form a Wheatstone bridge. The unbalance of the bridge is a measure of the deformation of the gauges and hence of the force applied to them. Alloys and semiconductors are widely used as gauge materials.

In conventional strain-gauge transducers a resistive foil is attached to a thin plastic sheet that is bonded by adhesive to the elastic element to be subjected to the strain. New types of thin-film strain gauges were first used in pressure transducers in the early nineteen sixties [3]. The thin films are deposited directly on an elastic element by evaporation or sputtering. An example of an experimental thin-film pressure transducer is shown in *fig. 1*. An attraction of using thin-film technology for such devices is that it can be used in mass production, with considerable cost reduction [4]. For the transducers it means improved quality and a substantial increase in the range of applications. We have succeeded in making thin-film strain gauges that are useful not only for conventional applications but also for measurements under special conditions, such as high and low temperatures, rapid temperature changes, high humidity, corrosive environment and severe vibration. In these thin-film strain gauges the resistive layer consists of an alloy. In passing we should note that it is also possible to make thin-film strain gauges in which the resistive material is a polycrystalline semiconductor, giving an increase

Dr K. Bethe and Dr D. Schön are with Philips GmbH Forschungslaboratorium Hamburg, Hamburg, West Germany.

Table I. Advantages of strain gauges compared with inductive and capacitive sensors.

Technical advantages	Economic advantages in manufacture
Very good linearity	Few manual operations
Low temperature sensitivity	Low rejection rate
Very small mass	No mechanical alignment
No moving parts	Very few individual parts
Easily protected against aggressive environment	Low costs
Local measurement	

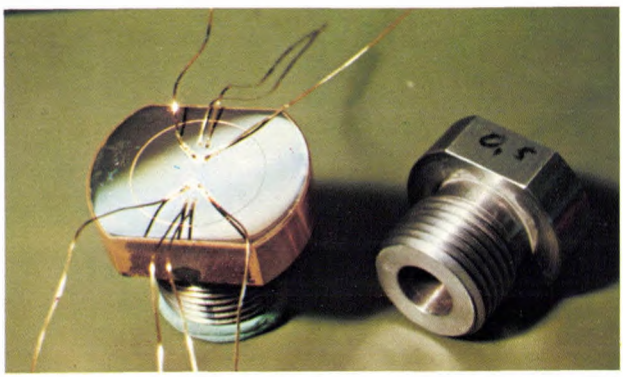


Fig. 1. Experimental pressure transducer with thin-film strain-gauge system, seen from two sides. This transducer was designed for measuring pressures up to 70 bars.

of about ten times in sensitivity. Although these gauges have a limited linearity and their resistance is rather temperature-dependent, they can be quite useful for simple applications.

In this article we shall first present some results of our investigation of the stages in the process and the precautions required in producing thin-film strain gauges that have the desired characteristics. We shall then discuss some elastic elements that we have designed for these gauges, and finally we shall compare the characteristics of thin-film strain gauges with those of conventional strain gauges.

Technology of thin-film strain gauges

The various stages in the manufacture of thin-film strain gauges are summarized in *Table II*. We shall now examine each of these stages individually, confining ourselves to gauges produced by means of r.f. sputtering in argon [5].

Pretreatment of the substrate

Formerly the substrate surface on which the thin films were to be deposited was usually polished with an abrasive, e.g. diamond powder. This method has been replaced by a mechanical-chemical treatment using silicon dioxide in an alkaline suspension. This technique is only applicable to corrosion-resistant materials, however, such as stainless steel and some copper alloys. We decided to use the more advanced technique of electropolishing, which is faster, easier and therefore cheaper, and can be used for high-homogeneity steels with a very low carbon content and for beryllium copper.

Deposition of the thin films

Fig. 2 is a diagram representing the thin-film structure of our strain gauges. The substrate and the resistive film are separated by an electrically insulating metal-oxide film. To guarantee the adhesion of this insulating film to the substrate it is first of all necessary to remove as many of the water, metal-oxide or hydroxide molecules adsorbed on it as possible. Also, the insulating film must somehow or other be mechanically and chemically bonded to the substrate. With a steel substrate we generally use a 0.1 μm thick adhesion layer consisting of a ternary nickel-chromium alloy, the same material that will be used in a later stage for the resistive film. The adhesion is increased by the relatively high energy of the sputtered particles and by the cleaning effect of the glow discharge on the surface of the substrate.

The insulating oxide film has to be sputtered with a carefully controlled oxygen content to ensure that the film has the desired stoichiometric composition. In this way insulation resistances of more than 10^9 Ω can easily be achieved. Instead of using a single oxide film for the insulation we use a sandwich structure, with a layer of magnesium oxide between aluminium oxide, because such a structure contains far fewer pinholes for the same total thickness (about 2 μm). At the interfaces between the insulating materials the vertical pore propagation is interrupted, owing to altered nucleation conditions. In addition, the formation of spinel ($MgAl_2O_4$) at the interfaces could contribute to structural repair. Another advantage of this multiple-layer insulation is that it allows a relatively large maximum strain ($> 10^{-2}$). The probable reason for this is that cracks at the interfaces between the different materials cannot propagate further. The use of magnesium oxide and aluminium oxide for the insulating films has further advantages. The thermal

Table II. Process stages in thin-film strain-gauge technology.

1. Pretreatment of substrate
2. Thin-film deposition in a vacuum unit
3. Formation of pattern for Wheatstone bridge
4. Stabilization and adjustment of temperature coefficient of resistance
5. Trimming of bridge circuit for symmetry
6. Formation of electrical connections
7. Deposition of passivating film

[1] See for example T. Potma, Strain gauges, Centrex, Eindhoven 1967, and A. C. M. Gieles and G. H. J. Somers, Philips tech. Rev. **33**, 14, 1973.

[2] Transducers giving a large output signal have lost some of their attraction now that cheap IC amplifiers are available. They are still very useful, however, in applications where there is considerable electrical noise, as in cars.

[3] P. R. Perino, Instr. Control Syst. **38**, Dec. 1965, p. 119.
R. L. Cheney, Industr. Res. **18**, April 1976, p. 63.
W. Ort, Messtechn. Briefe Hottinger Baldwin Messtechnik **13**, 7, 1977.

[4] K. Bethe and D. Schön, in: Philips — Unsere Forschung in Deutschland, Vol. III, p. 265, 1980.

[5] We reported on this at the 7th Conf. of the IMEKO technical committee on measurement of force and mass, Braunschweig 1978.

expansion coefficient of these materials is close to that of the substrate and they have a high thermal conductivity, a high melting point and relatively high optical transmission (transparency). These properties are especially important for trimming and the subsequent production of patterns by means of a laser.

After the oxide film has been formed a resistive film 0.3 μm thick is sputtered. The ternary alloys CrNiM, in which M represents Mo, Ti, Al or Si are very suitable for this film [6]. Films produced with these alloys have a relatively high resistance and are very stable. They also have an almost zero temperature coefficient of virtual strain. On the other hand, film properties such as the temperature coefficient of resistance, creep and strain sensitivity are highly dependent on the sputtering conditions. Some important sputtering variables are substrate temperature, argon pressure, residual gas (nitrogen, oxygen, water vapour), r.f. power, substrate and target potential and the geometry of the glow discharge. The associated microscopic effects consist of variations in crystallite size, texture, phase equilibria and in the composition of the films, due for example to included inert gases, chemically bonded reactive gases and preferential resputtering of the component that sputters most easily.

The final film to be sputtered is a 1 μm gold layer for contact areas and internal connections.

Pattern formation

The bridge pattern in the resistive film and the bonding pads in the gold film are made in two photolithographic stages. The etching operations can be either chemical (wet etching) or sputter etching (back-sputtering). To obtain good results with wet etching, moderate deposition temperatures are necessary when sputtering the gold film. Otherwise, diffusing components in the resistive film would cause the formation of an etch-resistant interface, giving rise to undercutting. In contrast, the more expensive back-sputtering process does not cause undercutting and removes any gold contamination of the resistive elements.

It is easier and cheaper to produce the patterns at the same stage as the deposition by using a shadow mask. This method has a number of disadvantages, however. A mechanical mask cannot be made as accurately as a photomask, which can have a resolution of 1 μm. A mechanical mask is restricted to very simple structures, and it is subject to dimensional changes during use as a result of deposits of material at its edges. The edges are poorly defined since some of the material scatters under the mask. This effect is particularly pronounced in sputtering deposition because the mask buckles in places, owing to the high energy of the particles that strike its surface. These disadvantages make the mask technique

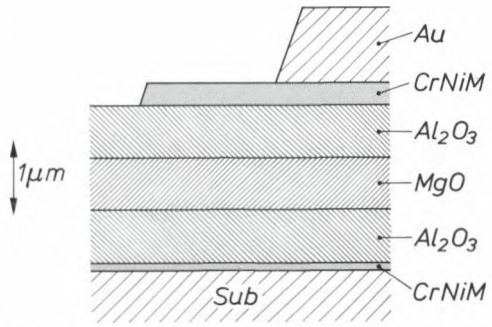


Fig. 2. Diagram of cross-section of a thin-film strain-gauge transducer. Sequence of layers on substrate *Sub*: adhesion layer CrNiM, insulation sandwich of $Al_2O_3/MgO/Al_2O_3$, resistive layer of CrNiM with pattern for Wheatstone bridge, and gold layer with pattern for electrical connections.

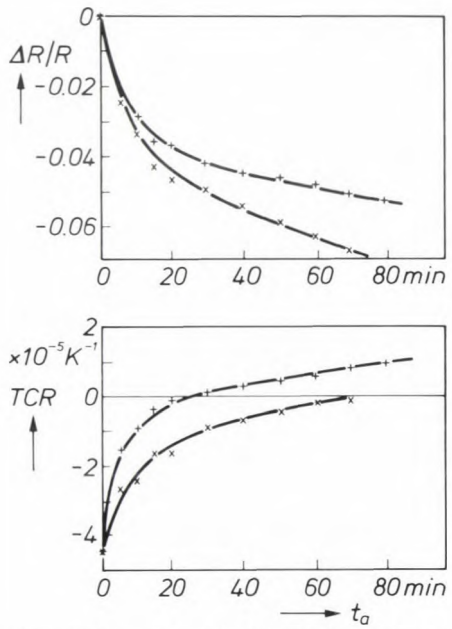


Fig. 3. Relative change in resistance $\Delta R/R$ and the temperature coefficient of resistance (TCR) for two resistive films, as a function of annealing time t_a at 300 °C. The differences between these two layers are due to a different substrate temperature during sputtering. By using the annealing time as a process variable, temperature effects can be made negligibly small (TCR ≈ 0).

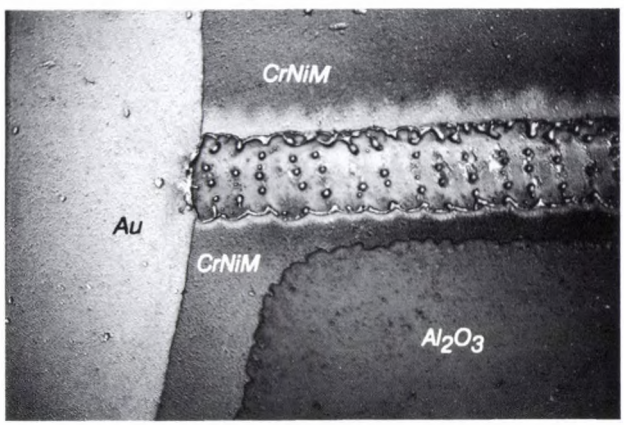


Fig. 4. Trim cut of 20 μm width in a resistive film, made by a frequency-doubled pulsed YAG:Nd laser, emitting at 530 nm with a pulse length of 100 ns.

[6] J. H. Mooij, Phys. Stat. sol. (a) **17**, 521, 1973.

unsuitable for producing resistive elements; it can however be used for the less critical process of forming the bonding pads with typical dimensions of 0.5×0.5 mm.

Additional stages

The temperature coefficient of resistance (TCR) of the resistive elements thus obtained usually has an undesirably low negative value. This can be altered by a thermal after-treatment of the elements at 300 °C. From the ageing diagrams for two different elements in *fig. 3* it can be seen that the electrical resistance decreases whereas the value of TCR increases as the annealing continues. By choosing the appropriate annealing time the influence of temperature on the resistance can be made negligibly small.

A Wheatstone bridge has to be 'trimmed' for symmetry, and we do this by means of a numerically controlled laser. Strain-gauge material is burnt away from the resistive film until the bridge is balanced. Broad trimming areas are incorporated in each separate strain gauge for this purpose beforehand. A typical laser cut is shown in *fig. 4*.

The electrical connections of the strain-gauge bridge are formed by four gold leads that are bonded to the bond pads by thermocompression.

The final stage is the deposition of a passivating film. While this has to be sufficiently impermeable to water, it must not affect the elastic behaviour of the elastic element; this means that the film thickness is limited to about 10 μm.

The thin-film process has been optimized step by step for maximum permissible strain, load changes, creep, zero drift, long-term stability and corrosion resistance. *Fig. 5* shows a Wheatstone bridge we have made with thin-film strain gauges, deposited on a thin rectangular bending 'beam' or bar of stainless steel. The dimensions of the resistive elements are 3×0.1 mm. Data for this bridge are given in *Table III*.

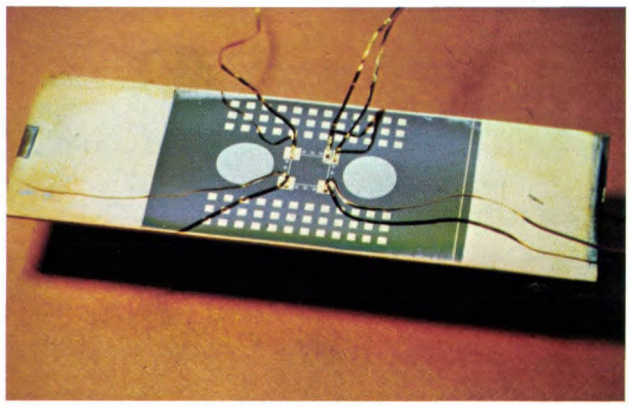


Fig. 5. Thin-film strain gauges forming a Wheatstone bridge with two active gauges and two passive gauges as dummy arms. The gauges (3×0.1 mm) have been deposited on a thin stainless-steel sheet with dimensions of 75×25×1 mm. Also shown are areas for testing insulation and resistance layers and the quality of the photolithographic patterns.

Elastic elements for thin-film strain-gauge transducers

The elastic element that converts the physical quantity to be measured (e.g. a force) into a strain has to meet a number of general requirements. The conversion should be linear, without hysteresis and independent of time. It should be electrically as well as mechanically linear. This implies that each gauge in the Wheatstone bridge must have the same absolute value of strain. A large strain is necessary to give a strong output signal. The elastic element must be insensitive to external disturbances and to temperature variations or gradients, while creep and hysteresis should be as small as possible.

In addition to these general requirements, the use of thin-film strain gauges imposes further restrictions on the choice of the elastic material and the design of the elastic element.

Choice of elastic material

The elastic material must be capable of withstanding the conditions prevailing during the various stages in the thin-film technology. For instance, its elastic properties must not be impaired by the heat treatment at 300 °C and the substrate surface must be easily polished. A very important requirement for the elastic material relates to its homogeneity. The area on which the thin films are deposited must be free of pores and non-metallic inclusions, otherwise pinholes might form in the insulation layer and result in a poor or

Table III. Typical data for our thin-film strain gauges.

Surface resistivity	5 . . . 16 [*] . . . 30 Ω/□
Resistance of single gauge	50 . . . 500 [*] . . . 10 000 Ω
Temperature coefficient of single gauge	$\pm 2\times10^{-6}$ K^{-1}
Temperature coefficient of bridge zero (zero drift)	$\pm 3\times10^{-4}$ K^{-1} [**]
Temperature coefficient of sensitivity	$+ 3\times10^{-4}$ K^{-1} [**]
Strain sensitivity [***]	1.9
Creep	$< 10^{-4}$ [**]
Resistance of insulation layer	$> 10^{9}$ Ω
Maximum static strain	10^{-2}
Number of permitted load reversals at strain of 10^{-3}	$> 10^{6}$

[*] Standard value.
[**] Value with respect to full-scale reading at strain of 10^{-3}.
[***] Ratio of relative change in resistance to strain.

unstable insulation resistance as well as an unacceptably low breakdown voltage between strain gauge and elastic element. This means that elastic materials can only be successfully combined with thin-film strain gauges if they have been previously refined by an additional remelting process. This is preferably done in an electric arc furnace under a molten reactive slag, which is very effective in reducing oxides and sulphides in steels.

Low creep is a vital property of elastic material for a transducer with thin-film strain gauges, since there is no possibility of creep compensation in these devices. Within the transducer area the strain of the elastic element will not exceed 2×10^{-3}. At this strain amplitude the observed creep is due entirely to diffusion, whereas there is no intercrystalline slip; this is verified by the absence of hysteresis. We have established this by measuring the time dependence of after-effects. When measuring creep in elastic materials it is necessary to distinguish between the viscous after-effect due to diffusion and the adiabatic temperature change (about 0.1 °C) that causes a time-dependent change of dimensions as well. Our measurements have shown that the magnitude and time constant of non-elastic after-effects can vary by more than a factor of 10. Thus, the creep with respect to a full-scale reading at a strain of 10^{-3} is found to be 13×10^{-4} in a high-strength aluminium alloy, but for maraging steel it is only 0.8×10^{-4}. In general, precipitation-hardened materials seem to be superior to those hardened by a transformation reaction.

Taking into account elastic behaviour, thermal conductivity, corrosion resistance, homogeneity, ease of machining, availability and price, we chose two precipitation-hardened materials: stainless steel with 17% of chromium, 4% of nickel and 3% of copper, and beryllium copper (98% of Cu, 2% of Be).

Design of elastic element

For an elastic element the use of thin-film strain gauges has two fundamental pecularities. Firstly, the entire elastic element has to be contained inside the vacuum chamber for deposition of the thin films. Secondly the deposition itself is characterized by a transport of material in the direction normal to the target surface. In conventional sputtering equipment with planar target, only plane surfaces can be coated with the necessary uniformity of thickness and composition. These conditions for the thin-film process impose further economic requirements: the elastic element should be as small as possible and all the strain gauges for the Wheatstone bridge should be positioned on a single flat surface, whose dimensions should be small to guarantee optimum uniformity.

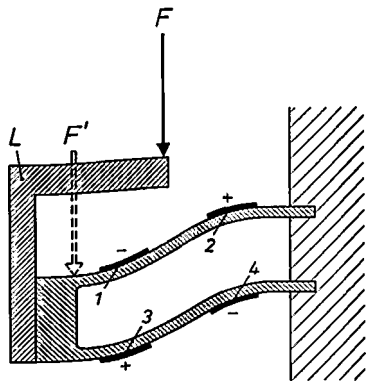


Fig. 6. Diagram representing a double bending bar, clamped at one end. The strain gauges *1*, *2*, *3* and *4* form a Wheatstone bridge. A force can be applied either directly (arrow *F'*) or, preferably, via a torque-compensating load lever *L* (arrow *F*). When the force is applied, gauges *1* and *4* have a negative strain, and gauges *2* and *3* have a positive strain.

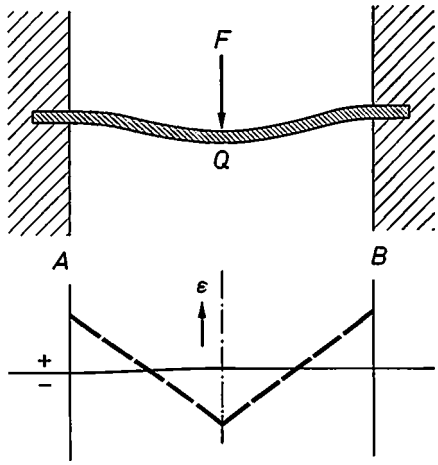


Fig. 7. *Above:* Diagram of cross-section of a symmetrical bending bar, clamped at both ends *A* and *B* and subjected to a force *F* at the centre *Q*. *Below:* Variation of the strain ε along the beam.

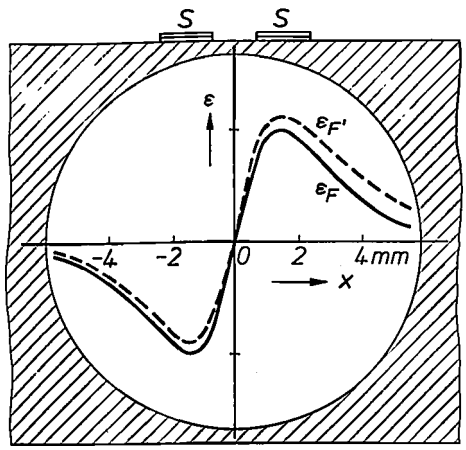


Fig. 8. Diagram of the deformation regions of a single-hole double bending bar. The strain is calculated as a function of position *x* along the bar. Without the load lever of fig. 6 the strain ($\varepsilon_{F'}$) is not symmetrically distributed. With the load lever the strain (ε_F) is symmetrically distributed, as confirmed by experimental strain measurements. The best positions for the strain gauges *S* for the Wheatstone bridge are indicated, two of them at places with maximum positive strain, the other two at places with maximum negative strain.

As these limiting factors do not apply to conventional transducers their elastic elements usually do not meet the above requirements of thin-film technology. It was therefore necessary to design elastic elements specially suited to the application of thin-film strain gauges. We shall now describe two force transducers that we have designed from known principles and which do meet the requirements.

Fig. 6 shows a diagram of the construction of a double bending bar. In principle, this bar can be considered to be derived from the familiar symmetrical bar clamped at both ends and loaded at its centre; see *fig. 7*. Cutting the symmetrical bar at its centre Q and rotating the left-hand half AQ under the right-hand half QB results in the double bending bar shown in fig. 6, if both halves are rigidly connected at Q. Complete correspondence with the symmetrical bar clamped at both ends is achieved by applying the force via a bent load-lever that compensates for the residual parasitic bending moment. Starting from this widely used design we arrive at the desired compact device by introducing a locally variable cross-section. This is done by drilling a circular hole through the width of the bar, thus introducing a variation in the moment of inertia that results in a strain distribution as shown in *fig. 8*. This distribution is symmetrical and the positions of the maximum and minimum are close together. Outside these deformation regions of only a few millimetres the strain quickly diminishes as a function of the distance from the centre. This concentration of strain compares favourably with the original strain distribution of the symmetrical bending bar (fig. 7). In this compact design there are fewer unwanted effects that interfere with the force to be measured and the frequency response is better.

Force transducers of this single-hole double bending-bar type can be used for loads in the range from 10 N to 5000 N. We used beryllium copper as elastic material for the lower loads and stainless steel for the higher loads. *Fig. 9* shows two transducers for maximum loads of 200 N and 2500 N. Both transducers carry a complete Wheatstone bridge on the upper side. When positioning the strain gauges on one side it is necessary to compensate for the parasitic bending moment by a load lever to avoid a tilt of the free end. The result of omitting this lever is considerable nonlinearity, which can be as much as 1.0×10^{-3} of the full-scale reading. This compares with a maximum nonlinearity of only 1.5×10^{-4} when strain gauges are located on both sides or when a load lever is used; see *fig. 10*.

Another very compact elastic element for force transducers (load cells) is a centrally loaded circular plate clamped at the edge, which may be considered as the axially symmetrical analogue of the symmetrical bending bar in fig. 7. Making the centre rigid reduces the inherent nonlinearity to an acceptable value [7]. Most problems usually encountered with elastic transducer elements of this type originate from poor stability of the flange; the effect of this is that the strain is not linear, with hysteresis, and is not reproducible. This problem can be solved, however, by using two circular plates instead of one, bolted to-

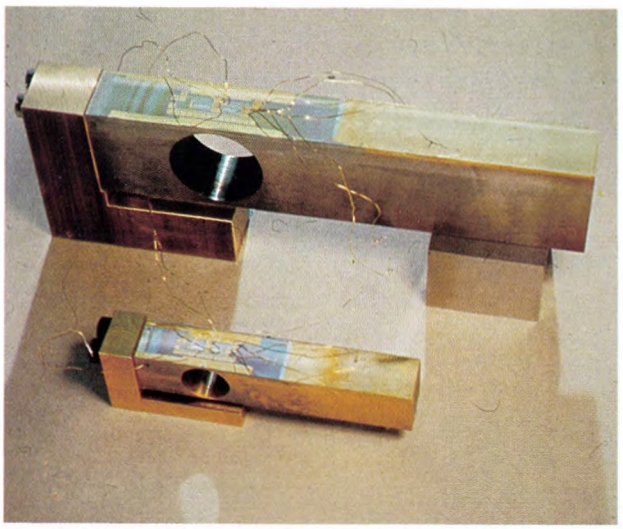


Fig. 9. Two force transducers with a single-hole double bending bar, fitted with thin-film strain gauges. The small transducer can be used for forces up to 200 N, the large one for forces up to 2500 N.

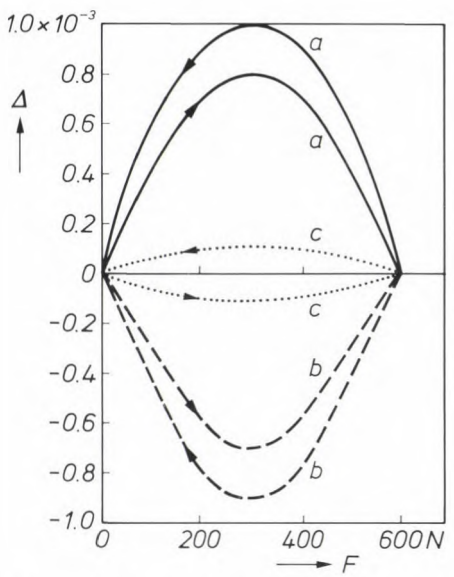


Fig. 10. Linearity deviation Δ as a function of the applied force F for a single-hole double bending bar. The arrows indicate whether the load increases or decreases. Curves *a* and *b* were obtained with no load lever. In *a* all four strain gauges were on the upper side of the bar, in *b* they were all on the lower side. The deviation from linearity is considerably less when a load lever is used, or when two strain gauges are on the upper side and two on the lower side (curve *c*).

[7] K. Bethe and D. Schön, 6th Conf. of the IMEKO technical committee on measurement of force and mass, Odessa 1977, p. 48.

gether face to face at the outer flanges; see *fig. 11*. By means of suitable profiling the strain here too is concentrated in very narrow sensing areas. This design of circular twin load cell can be used for measuring forces in the range from 1 kN to 100 kN. *Fig. 12* shows a cell we have produced for loads up to 30 kN. Its maximum deviation from linearity (total error) is 1.5×10^{-4} of full-scale reading.

Table IV. Advantages of thin-film strain gauges.

Maximum working temperature +250 °C [*]
Lowest working temperature −250 °C
Very good thermal coupling
Very low hysteresis and creep
High stability
Insensitive to humidity
Negligible mechanical stiffness
Easily miniaturized
High resistance possible, even with very small dimensions
Exact location on elastic element
Integration of additional elements for compensation of temperature sensitivity
Automatic trimming by laser
Low tolerances
Substantial reduction of costs in mass production

[*] Even higher temperatures are possible when alloys such as platinum-tungsten are used.

Table V. Advantages of conventional strain gauges.

Positive creep of elastic element can be compensated by negative creep of plastic film
High breakdown voltage between strain gauge and elastic element
Insensitive to local inhomogeneities of elastic material
Not affected by surface quality of elastic element
Structure of elastic element not predetermined
Elastic element not limited in size
No economic limit for minimum quantity

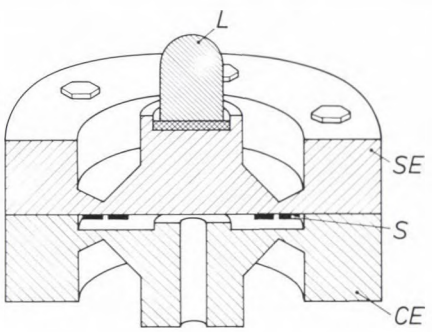


Fig. 11. Diagram of cross-section of a twin load cell. The sensing elastic element *SE* consists of a profiled circular plate, with the strain gauges *S* applied to the underside. A compensating element *CE* is formed by a second circular plate with same profile. The outer flanges of the plates are bolted together. *L* load button.

Fig. 12. Elastic element of a circular twin cell for measuring forces up to 30 kN. The element is covered with a large number of thin-film strain gauges. The outer diameter of the circular plate is 70 mm.

Comparison of conventional and thin-film strain gauges

The thin-film strain gauge does not enter a new market; it has an established competitor in the highly developed conventional strain gauge. Comparison of the two types of strain gauge has a number of technical and economic aspects. As can be seen from *Table IV* and *Table V*, both alternatives for strain sensing have their own specific field of applications resulting from their different features, although there is some overlap for medium-sized transducers.

The use that will be made of thin-film strain gauges is highly dependent on the demand for large quantities of transducers manufactured in the same thin-film process. This can include different types, however, if the differences in dimensions of the elastic element are not relevant in thin-film processing, e.g. the thickness of the bending bar.

Thin-film strain gauges are particularly suitable for small transducers, permitting miniaturization that would be impossible with conventional strain gauges. On the other hand the loading of thin-film strain-gauge transducers is subject to a kind of economic upper limit, which will be about 50 kN. In special cases, however, there may be advantages in using thin-film strain-gauge technology even at higher loads.

For quantity production the thin-film process offers a marked cost reduction, provided that the active transducer area is small, i.e. less than 10 cm^2. *Fig. 13* shows a number of inexpensive sensors for measuring displacement or acceleration. Each sensor consists of a triangular bending bar of beryllium copper and a semi-active Wheatstone bridge with two active arms and two passive gauges as dummy arms; the coated area is about 2 cm^2. Batch processing on a multiple

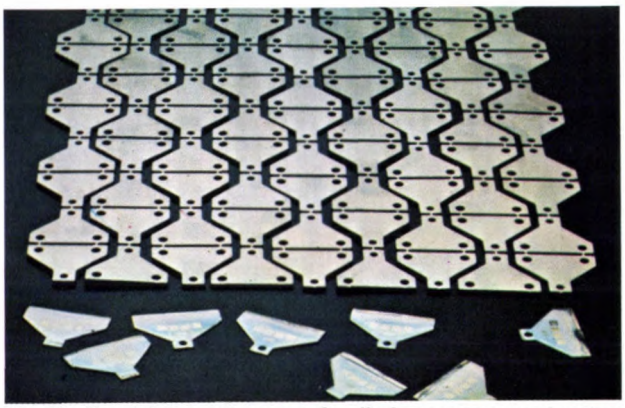


Fig. 13. *Front:* Low-cost sensors for displacement or acceleration measurements. The elastic elements are formed by triangular bending bars made from etched beryllium-copper sheet. Each element is coated with a Wheatstone bridge with two active and two passive thin-film strain gauges and trimming areas. *Back:* Multiple frame of beryllium copper used for simultaneous production of a large number of these sensors.

beryllium-copper frame permits a large number of sensors to be made simultaneously. This can reduce costs by a factor of three.

We have shown that thin-film strain gauges are a valuable addition to the range of transducers for the measurement of low loads and form an inexpensive alternative to other transducers for mass application. There is a range of overlap for transducers for forces between 5 kN and 50 kN, where there is no distinct preference for thin-film or conventional strain gauges. In this range the final choice will depend on the particular requirements (temperature, humidity, etc.) and on the number of transducers required.

Summary. Strain-gauge transducers can be used for measuring mechanical quantities such as force and pressure. Its operation is based on a change in electrical resistance due to deformation of the gauge. The strain-gauge transducers described in this article contain a number of thin films deposited successively on an electropolished substrate by r.f. sputtering. The substrate is first coated with a very thin adhesion layer consisting of a ternary CrNi alloy. An insulating sandwich of $Al_2O_3/MgO/Al_2O_3$ is then deposited, followed by a resistive strain-sensing film formed from the same alloy as the adhesion layer. Finally a gold film is deposited, which is used for making electrical contacts. A Wheatstone bridge is formed in the resistive film by photolithography, and is trimmed for symmetry with a laser cutter. The use of thin-film strain gauges imposes difficult requirements on the choice of elastic material and the design of the elastic element. Stainless steel containing 17% of Cr, 4% of Ni and 3% of Cu, and beryllium copper (98% of Cu, 2% of Be) are found to be very suitable elastic materials. Preferred designs for the elastic element are a double bending bar with a cross-sectional hole for forces from 10 N to 5 kN, and a centrally loaded circular twin cell for forces from 1 kN to 100 kN. Thin-film strain-gauge transducers are particularly suitable for measuring small forces and for measurements under extreme conditions. They can provide a substantial cost reduction when mass produced.

An integrated echo canceller for baseband data transmission

P. J. van Gerwen, W. A. M. Snijders and N. A. M. Verhoeckx

Data transmission is an ideal field for the use of digital signal-processing methods. With the continuing advances in integrated-circuit technology — more transistors per unit chip area, lower power consumption per transistor, higher speed of operation — a corresponding increase in the quantity and complexity of the processing operations is now becoming available. In this article the authors describe a digital echo canceller that enables data signals to be transmitted simultaneously in both directions and in the same frequency band on a single pair of copper wires. This circuit, which contains some 12 000 transistors, only occupies a chip area of about 38 mm^2. By using this circuit expensive copper wire can be traded for inexpensive electronics.

Introduction

The telephone network has for many years been used for the exchange of information between a centrally located computer and remote peripheral units ('data transmission') [1]. A distinction can be made here between switched lines for communication between any two subscribers, and private lines ('leased lines'). In the not too distant future new switching and transmission systems specially designed for the transmission of data signals ('data networks') will be introduced. These will not only combine the particular advantages of the two types of existing data circuits where possible, but will also offer new facilities. A data network will have the following characteristic features:

— switched circuits
— rapid setting up of connections
— better transmission performance
— considerable variety in permitted transmission rates
— full-duplex operation, i.e. simultaneous transmission of signals in opposite directions
— efficient use of transmission media by multiplexing similar signals wherever possible.

There are some fundamental differences between a circuit in a data network (*fig. 1a*) and a switched or leased circuit in a telephone network (fig. 1*b* and *c*). In a data network the subscriber is only concerned with the transmission of his data by subscriber line from his own location to the nearest input of the central data network (*CDN*, fig. 1*a*). In the telephone network, on the other hand, a subscriber has to transmit his data signals in such a form that they can pass through the entire switching and transmission network without unacceptable distortion. This means that the subscriber has to have fairly complicated equipment for transmitting and receiving data ('voice-band modems'). In a data network it is possible to operate with a much simpler type of apparatus ('baseband modem') at each end of every subscriber line because such a line only consists of cables, which may be coil-loaded. Another advantage associated with this is the availability of a larger bandwidth, permitting a higher transmission rate than with an arbitrary circuit in the telephone network.

In the switched telephone network (fig. 1*b*) the subscriber lines and the outer parts of the switching and transmission network form a two-wire circuit that consists of a single pair of conductors on which it is in principle possible to transmit signals in both directions. The central part of the switching and transmission network forms a four-wire circuit on which the signals are transmitted separately in the two directions. The changeover from two-wire to four-wire circuit and vice versa is made in 'hybrids' (*Hyb*). The main purpose of the four-wire circuit is to permit signal amplification. It also allows signals that have to be transmitted over considerable distances in the same direction to be combined ('multiplexed') to form a single composite signal. This is often done by means of frequency-division multiplex in a carrier-telephone system (*Car*). A switched circuit is characterized by a non-ideal frequency characteristic with a passband from about 300 Hz to 3400 Hz. This is mainly because of the presence of transformers and loading coils in the subscriber lines, and filters in the amplifiers and multiplexing systems. In addition nonlinear effects may occur, and also frequency shifts and time-dependent phase variations ('phase jitter') as a result of modulation and demodulation in carrier systems. In spite of these shortcomings it is now possible to use voice-band modems to transmit data on switched lines at a maximum rate of 4800 bit/s. At any particular moment there is trans-

Ing. P. J. van Gerwen, Ing. W. A. M. Snijders and Ir N. A. M. Verhoeckx are with Philips Research Laboratories, Eindhoven.

mission in only one of the two directions ('half-duplex'). If it is desired to transmit data signals simultaneously in both directions, this can only be done by halving the maximum rate because of the imperfections in the hybrids. Another disadvantage of switched circuits is the time lost while a connection is being set up; this can be relatively long for short messages. In half-duplex traffic time is also lost when the transmission direction is reversed ('turn-around time'). Some of the above difficulties do not occur when a leased line is used. *Fixed* connections (*FC*) are then established in the telephone exchanges. If required, the frequency characteristic is improved and the two-wire circuit is replaced by a four-wire circuit.

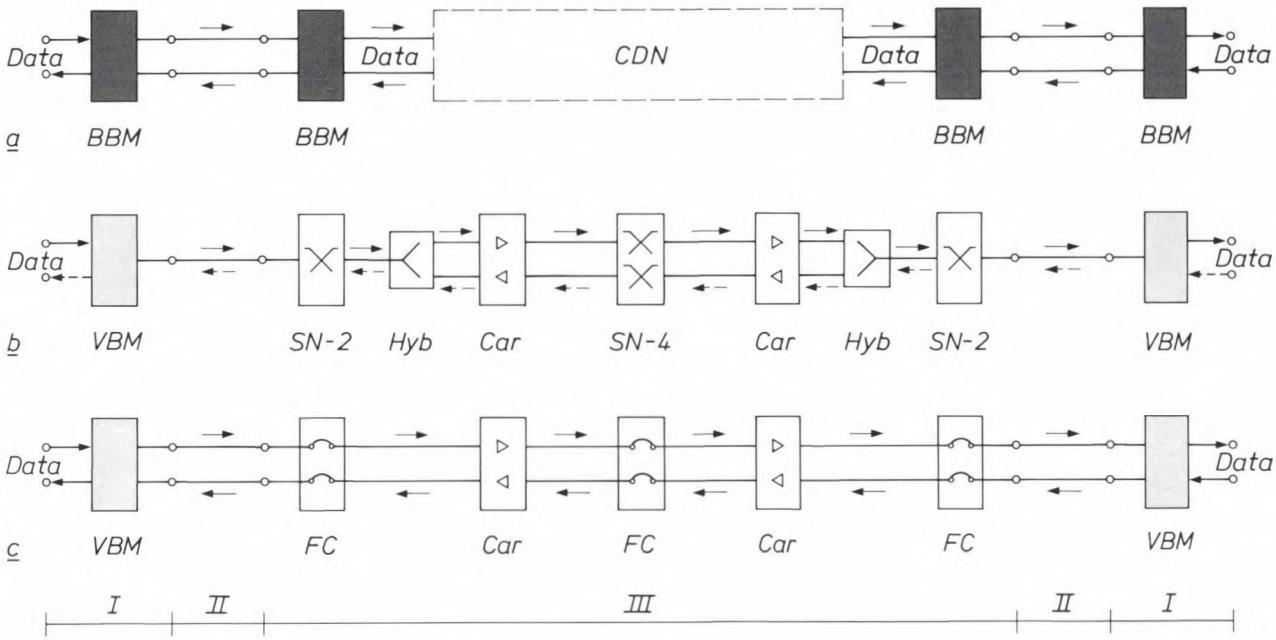


Fig. 1. Diagram showing three types of circuit used for data transmission: circuit in a data network (*a*), switched circuit in the telephone network (*b*) and leased circuit in the telephone network (*c*). In all three cases a distinction is made between subscriber's equipment (*I*), subscriber lines (*II*) and switching and transmission network (*III*). In connections made through the telephone network the subscriber has to use fairly complex voice-band modems (*VBM*) for the higher transmission rates, mainly because of the carrier systems (*Car*) that may be included in the transmission path. In data-network connections the subscriber can use a much simpler baseband modem (*BBM*) because he is only concerned with transmission to the nearest input of the central data network (*CDN*). A *single* solid connecting line represents a two-wire circuit; *two* solid lines represent a four-wire circuit. *SN-2* two-wire telephone exchange, *Hyb* two-wire/four-wire junction ('hybrid'), *SN-4* four-wire telephone exchange, *FC* fixed connection in exchange.

All this has the result that by using voice-band modems on leased lines it is possible to obtain a higher maximum transmission rate (at the moment 9600 bit/s) — even with full-duplex traffic — time is not wasted in establishing the connection and there is no question of 'turn-around time'. One obvious disadvantage is that a leased line does not have the versatility of a switched line. Finally, we should note that some leased lines in a telephone network are so short that the circuit only consists of a subscriber line. In such a case, baseband modems may also be used, just as for transmission in the outer part of a data network.

In this article we shall mainly be concerned with full-duplex traffic between two baseband modems and certain aspects closely associated with it. The commonest method of achieving full-duplex operation is to use a four-wire circuit, as shown in fig. 1*a*. It will be obvious that the number of cables required for the subscriber lines can be halved if a two-wire circuit could be used instead. This is possible if echo cancellers are used. The operation of existing circuits (the 'hybrids') used for two-wire/four-wire interfaces can then be improved to such an extent that they no longer form a limiting factor for full-duplex data traffic on a two-wire circuit. We shall be looking more closely into the basic principles of this shortly. In addition, we shall describe an integrated circuit (IC) for a digital adaptive filter [*]. This IC can be used in echo cancellers for baseband modems that work with various types of transmission code in common use.

[1] P. J. van Gerwen, The use of digital circuits in data transmission, Philips tech. Rev. **30**, 71-81, 1969.
F. de Jager and M. Christiaens, A fast automatic equalizer for data links, Philips tech. Rev. **37**, 10-24, 1977.
R. A. van Doorn and N. A. M. Verhoeckx, An I²L digital modulation stage for data transmission, Philips tech. Rev. **37**, 291-301, 1977.

[*] Important contributions to the design of this IC were made by J. J. W. Kalfs (Philips Research Laboratories, Eindhoven), and to the analysis of the problem by H. C. van den Elzen (Philips Video Division, formerly with Philips Research Laboratories, Eindhoven). We should also mention the co-operation with TRT (Paris) on echo cancellation.

Baseband modems

The signal-processing operations that have to be performed in a baseband modem (BBM) are determined by the characteristics of the transmission path (the subscriber line) and the nature of the information being transmitted (the data). The pairs of conductors that form the subscriber line have a usable frequency band that is limited at the top end by the attenuation alone, which increases with frequency and cable length. Baseband transmission is therefore possible at a large number of different bit rates. In practice, bit rates between 1.2 and 64 kbit/s are quite common. (However, if the conductor pair has loading coils, the frequency band is limited — to about 4 kHz — and only bit rates up to about 4.8 kbit/s can be used.) At the bottom end, the usable frequency band of the subscriber line does not extend to zero (d.c.), owing to the presence of transformers, which provide d.c. isolation. This means that the low-frequency components in the signal transmitted by the BBM should be very small, or zero. This can be arranged by coding [**] the data before transmission. In selecting the code other important factors have to be taken into account:

— At the receiving end there must be a simple way of synchronizing the receiver to the incoming signal (i.e. determining the clock frequency) independently of the information content of the transmitted data.

— Different kinds of signals being transmitted on different pairs in a cable must affect one another as little as possible (low cross-talk).

— The distance covered with a given maximum peak-to-peak value of the transmitted signal must be as large as possible.

— The equipment should be as simple (and hence inexpensive) as possible.

In practice there are a number of codes that meet these requirements, e.g. the differential biphase code, the Miller code and the AMI code. The first two of these give a two-valued coded signal and the third a three-valued coded signal. We shall return to these codes later.

Fig. 2a is a block diagram of a baseband modem and fig. 2*b* shows the signals at various points in this modem. The modem contains a transmitter *Tr*, a receiver *Rc* and clock circuits *Cl*. The data signal to be transmitted *1* is applied to the input of the transmitter. It is first processed in the scrambler *Scr* to break down periodic data patterns that would produce undesired discrete frequency components in the transmitted signal. With certain codes this also ensures that the transmitted signal always has sufficient clock information. The output signal *3* from the scrambler is also binary and is converted in a coder

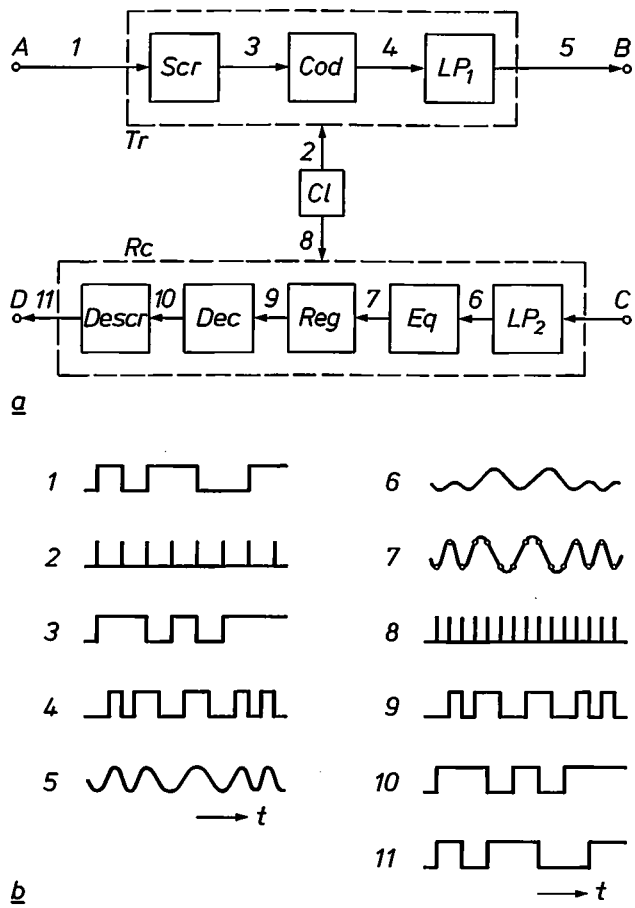


Fig. 2. Block diagram of a baseband modem (*a*) and the signals found in it (*b*). At *A* the data to be transmitted is presented to the transmitter unit *Tr* of the modem and is processed by a scrambler *Scr*, a coding circuit *Cod* and a lowpass filter LP_1 to make it suitable for transmission. After reception (*C*) the original data signal is recovered in the receiver *Rc* with the aid of a lowpass filter LP_2, an equalizer *Eq*, a regenerator *Reg*, a decoding circuit *Dec* and a descrambler *Descr*. The clock circuits *Cl* ensure that all the operations take place at the correct times. (For clarity we have chosen signals *5* and *6* such that they contain the same information, which means that signals *1* and *11* are identical.)

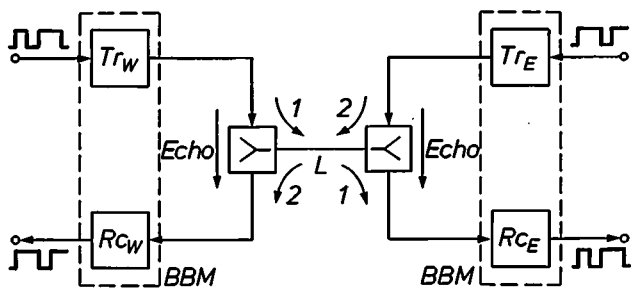


Fig. 3. Simultaneous transmission in two directions ('full duplex') between two baseband modems *BBM* on a two-wire circuit *L* by the use of hybrids. The arrows *1* indicate the path taken by the signal from the 'west' transmitter Tr_W to the 'east' receiver Rc_E. The arrows *2* relate to the signal path from the east transmitter to the west receiver. Since hybrid circuits operate far from ideally in practice, Rc_W receives a fraction of the signal transmitted by Tr_W — an 'echo' — which can seriously interfere with the reception of the desired signal. This also applies to Rc_E and Tr_E. The method illustrated here cannot therefore be used in practice.

Cod into a new binary or ternary signal *4*. After filtering (LP_1) this supplies a band-limited signal *5* that is suitable for transmission via the transmission path. In the receiver *Rc* the incoming signal is first filtered (LP_2) to suppress noise and interference as far as possible. The distortion due to the frequency response of the transmission path is corrected in an equalizer *Eq*. The resulting signal *7* can now be regenerated with the aid of the receiver clock *8*, whose frequency and phase are derived from the received signal. This gives a signal *9*, which corresponds to the coded signal *4* from the transmitter. Finally, the original data signal from the transmitter end is recovered with the aid of a decoding circuit *Dec* and a descrambler *Descr*.

The clock frequency in the transmitter is equal to the bit rate of the transmitted data. The same clock frequency must therefore be available both at the data source and at the transmitter. In principle, this frequency can be determined by making either the transmitter or the data source the 'master' and the other its 'slave'. In the receiver the clock frequency, or one of its harmonics, is derived from the incoming signal. The receiver must supply this clock frequency, as well as the data, at its output.

In data networks it is not unusual for the transmitter clock and the receiver clock to be coupled together. The bit rate is then exactly the same in both transmission directions. This is known as homochronous transmission.

Full-duplex transmission on two-wire circuits

There are various ways of achieving full-duplex transmission on a two-wire circuit between baseband modems. Two of the possible methods require a much greater bandwidth than with four-wire transmission. One is based on the use of different frequency bands in each of the two transmission directions (frequency-division multiplex or FDM). In the other methods data is transmitted alternately in each direction in 'bursts' at an increased rate. This is a form of time-division multiplex (TDM), which is also called the 'ping-pong' method.

At first sight, the ping-pong method, which we include here under full-duplex transmission, appears very like the description of half-duplex systems that we gave in the introduction (in the small print). The difference lies in the ratio between the rate at which the data is supplied by the source and the rate at which it is sent along the transmission path. In half-duplex operation this ratio is 1, and in the full-duplex ping-pong method the ratio is $\leq \frac{1}{2}$ (in practice it is often $\frac{1}{4}$). Although transmission never actually takes place simultaneously in the ping-pong method, the increase in rate has the effect that there seems to be a permanent connection available for the data sources in each direction.

In addition to these two methods there is a completely different method that requires no increase in bandwidth. The principle is illustrated in *fig. 3*. The figure shows a two-wire circuit L that has a four-wire baseband modem (fig. 2) connected to each end ('west' and 'east') by a hybrid circuit of the type used in two-wire/four-wire interfaces in the switched telephone network (fig. 1*b*). The function of the hybrid at the 'west' end of the line is to transfer the signal from the 'west' transmitter to the line and to route the incoming 'east' signal to the 'west' receiver. In practice, the hybrid, which in fact represents a bridge circuit, is not perfectly balanced and there is direct transfer ('leakage') from the west transmitter to the west receiver. In addition, irregularities in the cable cause a partial reflection of the west signal and this also arrives at the west receiver. These two effects result in an unwanted signal at the input to the west receiver, which we call an 'echo'. If the circuit is a long one and the hybrid balance is poor the echo at the input to the west receiver may be 30 dB above the received east signal. For the receiver to operate satisfactorily, however, this ratio must be considerably smaller (e.g. −20 dB). The method illustrated in fig. 3 is therefore completely unusable unless extra precautions are taken. (Similar arguments also apply to the other end of the circuit, of course.)

At first sight it may seem strange that the same kind of hybrid circuit that is widely used as the four-wire/two-wire interface for speech signals in a telephone network is so unsuitable for data signals. The explanation is that although echoes are also encountered with speech signals, they only become a nuisance if there is a relatively large delay, as in satellite communications. For a data receiver, however, only the level of the echo is important.

As we said in the introduction, this problem can be solved by eliminating the echoes at the input of the receiver by using an adaptive filter. Designing the adaptive filter in digital form can give the required 50 dB reduction in the echo in the above example, and an integrated-circuit version of the filter requires little space and takes little power. In the following sections we shall examine the various aspects of this echo cancellation more closely.

Echo cancellation

The echo signal that occurs when a hybrid circuit is used can be considered as the output signal of an undesired signal path (the 'echo path') between the

[**] For the adaptation of the spectrum of a signal to the frequency characteristic of a transmission path both of the terms 'coding' and 'modulation' are used. The term 'modulation' is generally preferred where a carrier signal is used, but otherwise no sharp distinction is made. The names Miller code and 'delay' modulation, for example, really refer to the same operation. In baseband data transmission it is usual to speak of (de)coding and not of (de)modulation, so that the use of the term 'baseband modem' is somewhat inconsistent ('modem' is a contraction of the words *mo*dulation and *demo*dulation). In our article, however, we have decided to follow common usage.

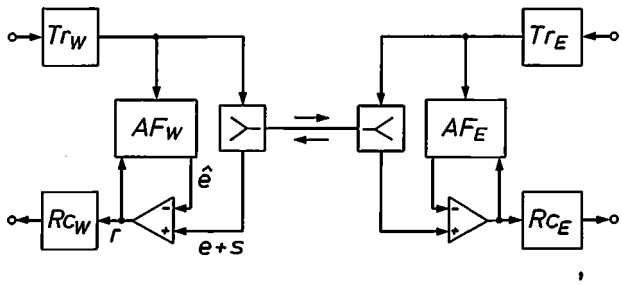


Fig. 4. The principle of full-duplex traffic on a two-wire circuit with hybrids and automatic adaptive echo cancellation. The echo signal e that travels from the west transmitter to the west receiver in a non-ideal hybrid is cancelled out by a simulated echo signal $\hat{e}$. The simulation is performed automatically by an adaptive filter AF_W that continously tries to minimize the correlation between the signal r and the transmitted west signal. When there is perfect cancellation ($e=\hat{e}$) the signal r is equal to the desired east signal s. There is a similar cancellation process in an echo canceller at the east end of the circuit.

lation is only obtained if the filter in use is as far as possible self-adjusting, i.e. is an automatic adaptive filter (*fig. 4*). Its setting is determined by an analysis of the signal that results after echo cancellation, since the correlation between this signal and the output signal of the transmitter provides a measure for the quality of the echo cancellation. In this process the filter makes use of a particular control algorithm that we shall discuss in more detail later.

In fig. 4, the echo signal at the west end is represented by e, the received east signal by s, the echo simulated by the adaptive filter (AF_W) is represented by $\hat{e}$, and r is the resulting signal after echo cancellation so that

$$r=s+e-\hat{e}. \qquad (1)$$

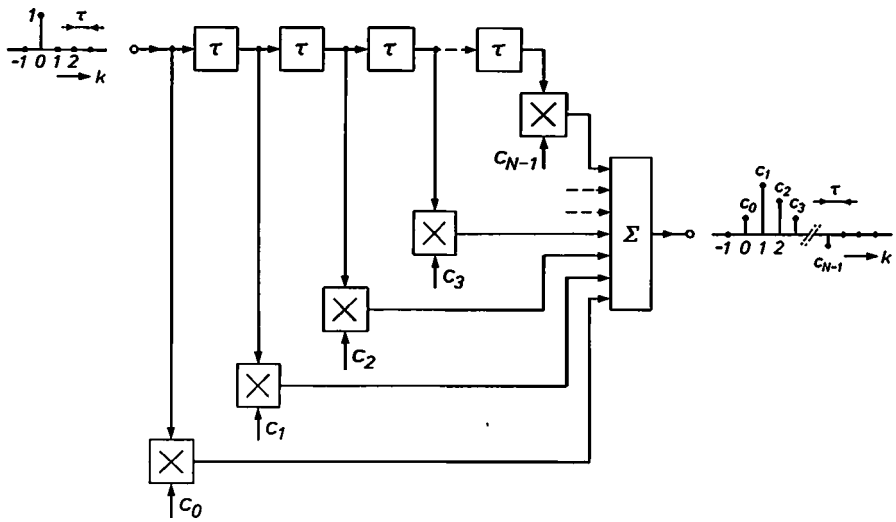


Fig. 5. Digital transversal filter. Successive samples from the input signal are stored in a series of delay elements with a delay τ, e.g. a shift register. For the calculation of a single output sample each of the stored samples is multiplied by one of the coefficients c_0, c_1, ... c_{N-1} and the products are added together. Because we are concerned here with a digital circuit, all the quantities such as input samples, coefficients and output samples occur as combinations ('words') of a given number of bits. In a graphical representation digital signals are usually indicated as shown here for the input and output signals of the filter. The input signal here is ...0,0,0,1,0,0,0,0,... (the digital unit pulse), so that the associated output signal (c_0, c_1, c_2, ... c_{N-1}) represents the impulse response of the filter. k is the discrete time variable.

transmitter and the receiver at the same end of the circuit. In baseband data transmission this echo path forms a linear transmission path that can only vary slowly with time (e.g. because of temperature fluctuations). It is therefore possible to generate a copy of the echo signal with the aid of a second signal path between transmitter and receiver in the form of a filter with the same transfer characteristic as the echo path. If the simulated echo is subtracted from the input signal of the receiver, the received echo is cancelled out. Because the echo path is not accurately known beforehand, may be different for each connection and may also vary a little with time, correct echo cancellation is only obtained if the filter in use is as far as possible self-adjusting, i.e. is an automatic adaptive filter (*fig. 4*). Its setting is determined by an analysis of the signal that results after echo cancellation, since the correlation between this signal and the output signal of the transmitter provides a measure for the quality of the echo cancellation. In this process the filter makes use of a particular control algorithm that we shall discuss in more detail later.

In fig. 4, the echo signal at the west end is represented by e, the received east signal by s, the echo simulated by the adaptive filter (AF_W) is represented by $\hat{e}$, and r is the resulting signal after echo cancellation so that

$$r=s+e-\hat{e}. \qquad (1)$$

The purpose of the echo cancellation is to make $\hat{e}$ as similar to e as possible. If the adaptive filter is perfectly adjusted ($e=\hat{e}$), r would in fact be equal to s.

Before describing the adaptive filter in more detail it is useful to pause for a moment to consider the input signal that is supplied to this filter. In fig. 4 the output signal from the transmitter is taken as the input to the filter. However, any other signal that produces the same output signal after linear processing can be used instead. This means, for example, that for the transmitter in fig. 2 we can use signal *4* instead of signal *5* as the input signal for the adaptive filter. The adaptive filter then automatically forms the product

of the transfer functions of LP_1 and of the echo path. The great advantage of this is that a binary digital signal is now fed to the adaptive filter instead of the analog output signal from the transmitter. This reduces the complexity of the adaptive filter, as we shall see, and it offers advantages in the control aspects. We shall not pursue this any further in the present article. All we need say here is that with most of the codes now used for baseband data transmission it is possible to obtain a binary digital signal from the transmitter that can be used as the input signal for the adaptive filter.

The adaptive digital filter

In the preceding sections of this article we have already mentioned some of the reasons why it is an advantage to use a digital form for the adaptive filter. A digital filter of the transversal type [2] is particularly suitable because these filters can provide a very good approximation to the impulse response of the echo path.

The principle of such a filter is illustrated in *fig. 5*. It consists of a series of delay elements of delay τ, in which successive values ('samples') of the digital input signal are stored, a number of multipliers in which the delayed samples are multiplied by the coefficients $c_0, c_1, \ldots c_{N-1}$ and an adder in which the products are added together. The impulse response of the filter, and hence its frequency characteristic, are determined by the values of the coefficients $c_0, c_1, \ldots c_{N-1}$, which are stored in a digital memory. The frequency characteristic of a transversal filter is periodic in $1/\tau$ and can only be freely specified in the frequency range from 0 to $1/2\tau$ Hz. For example, if a filter of this type is to be used to filter a data signal with a symbol interval of T seconds and τ is made equal to T, then the frequency characteristic of the filter can only be freely selected in the frequency range from 0 to $1/2T$ Hz. As in other applications, this is also frequently insufficient for filters for echo cancellation and τ is made equal to T/K (where K is an integer), giving a basic frequency range from 0 to $K/2T$ Hz. We then speak of an interpolating filter with an interpolation factor K, because the output samples of the filter appear at a repetition rate K times that of the original data symbols. However, input samples also have to be presented at time intervals τ to the input of the filter in fig. 5. Starting from an original data signal in which the samples (data) occur at time intervals T, a new digital signal with an interval between the samples of $\tau = T/K$ must first be derived. This is usually done by inserting $K-1$ samples with the value 0 between each pair of original samples.

By making the coefficients of a digital transversal filter variable and continuously adapting their values by using an appropriate algorithm an *adaptive* filter is produced. The coefficients are calculated in a control circuit that receives signals from which it is possible to determine how accurately the coefficients have been set. In the case of echo cancellation the data signal originally transmitted and the signal resulting after echo cancellation can be used for this; see *fig. 6*. To show that a closed loop is created, the subtracter in which the simulated echo is subtracted from the incoming signal is also shown in the figure. Since we are dealing with a digital system, in which only the values of the various quantities at discrete times k are important, all the signals and the variable coefficients are indicated as a function of k. For each value of k the adaptive filter calculates a sample of the simulated echo $\hat{e}(k)$ from the stored data $a(k)$, $a(k-1), \ldots a(k-N+1)$ and the instantaneous coefficients $c_0(k)$, $c_1(k), \ldots c_{N-1}(k)$ by means of the relation

$$\hat{e}(k) = \sum_{i=0}^{N-1} c_i(k) a(k-i). \qquad (2)$$

In addition a new value for each coefficient is calculated from

$$c_i(k+1) = c_i(k) + \Delta c_i(k), \quad i = 0, 1, \ldots N-1, \qquad (3)$$

where $\Delta c_i(k)$ is a function of the stored data $a(k)$, $a(k-1), \ldots a(k-N+1)$ and the value of the 'residual echo' signal $r(k) = s(k) + e(k) - \hat{e}(k)$. The exact relationship between $\Delta c_i(k)$ on the one hand, and the stored data and the residual echo on the other, is determined by the algorithm employed by the adaptive filter. For echo cancellation in data transmission the most suitable algorithm is the 'stochastic-iteration algorithm', for which

$$\Delta c_i(k) = 2\alpha\, r(k)\, a(k-i), \qquad (4)$$

where α is a constant. Combining equations (3) and (4) gives:

$$c_i(k+1) = c_i(k) + 2\alpha r(k) a(k-i), \quad i = 0, 1, \ldots N-1. \qquad (5)$$

From this equation it can be shown that, as long as there is correlation between $r(k)$ and the stored data, the coefficients are changed in steps (iteratively) such that this correlation is reduced and the echo cancellation therefore improved.

[2] The ideas on which transversal filters are based are due to H. E. Kallman, Proc. I.R.E. 28, 302, 1940. An explanation of the operation of such filters is also given by P. J. van Gerwen in his article of note [1].

As a measure of the equality of $\hat{e}(k)$ and $e(k)$ we take the mean square of their difference:

$$\varepsilon = \overline{(e(k) - \hat{e}(k))^2}.$$

The adaptive filter must be designed in such a way that it automatically makes the value of ε as small as possible. If we assume that $e(k) - \hat{e}(k)$ and $s(k)$ are statistically independent (which is very realistic since $e(k)$ and $\hat{e}(k)$ are dependent on the west data and $s(k)$ on the east data), then the minimization of ε is equivalent to the minimization of ϱ where ϱ is the mean-square value of $r(k)$:

$$\varrho = \overline{r^2(k)}.$$

Because $r(k)$ is equal to $s(k) + e(k) - \hat{e}(k)$ and $\hat{e}(k)$ is a linear function of each of the coefficients c_i (see equation (2)), it can easily be seen that ϱ is a quadratic function of each of the coefficients. The quantity ϱ is therefore at a minimum when

$$\frac{\partial \varrho}{\partial c_i} = \frac{\partial \overline{r^2(k)}}{\partial c_i} = 0, \qquad i = 0, 1, \ldots N-1.$$

In an adaptive filter, the coefficients can therefore be automatically adjusted to their optimum value (corresponding to a minimum value of ϱ) by means of the iterative algorithm:

$$c_i(k+1) = c_i(k) - \alpha \frac{\partial \overline{r^2(k)}}{\partial c_i}, \quad i = 0, 1, \ldots N-1$$

It has been found in practice that the algorithm works equally well if the mean value $\overline{r^2(k)}$ is replaced by the instanteneous value $r^2(k)$ provided that α is made sufficiently small. We then obtain:

$$c_i(k+1) = c_i(k) - \alpha \frac{\partial r^2(k)}{\partial c_i}, \quad i = 0, 1, \ldots N-1,$$

and using $r(k) = s(k) + e(k) - \hat{e}(k)$ and equation (2) this can be reduced to the stochastic-iteration algorithm given in equation (5).

By taking equation (5) as a basis it is possible to allocate a separate control circuit A_i to each of the coefficients $c_i(k)$ (*fig. 7a*). The resulting block diagram is shown in fig. 7*b*. The delayed input data $a(k-i)$ that has to be fed to the control circuit A_i can be taken from the delay line that is also used for the transversal filter.

The choice of the value of the control constant α in (5) is very important as can be seen from the following discussion. The behaviour of an echo canceller as a control system can best be explained with the aid of the ratio $P(e-\hat{e})/P(s)$ as a function of the time variable k. $P(e-\hat{e})$ represents the mean power (averaged over a short period of time) of $e(k) - \hat{e}(k)$ and $P(s)$ represents the mean power of $s(k)$. The ratio $P(e-\hat{e})/P(s)$ indicates the extent to which the signal $s(k)$ is disturbed by the residual echo $e(k) - \hat{e}(k)$. After a convergence period this ratio attains a constant value δ. In practice it is desirable that δ should be smaller than -15 dB. It can be shown that the control constant α has a direct influence on the value of δ: halving α reduces δ by 3 dB, giving better echo cancellation. But halving α also doubles the convergence time, which should be kept as small as possible in data transmission; see *fig. 8*. It is important therefore not to make α smaller than is necessary to obtain the desired value of δ. In practice this leads to values for α of 10^{-3} to 10^{-4}.

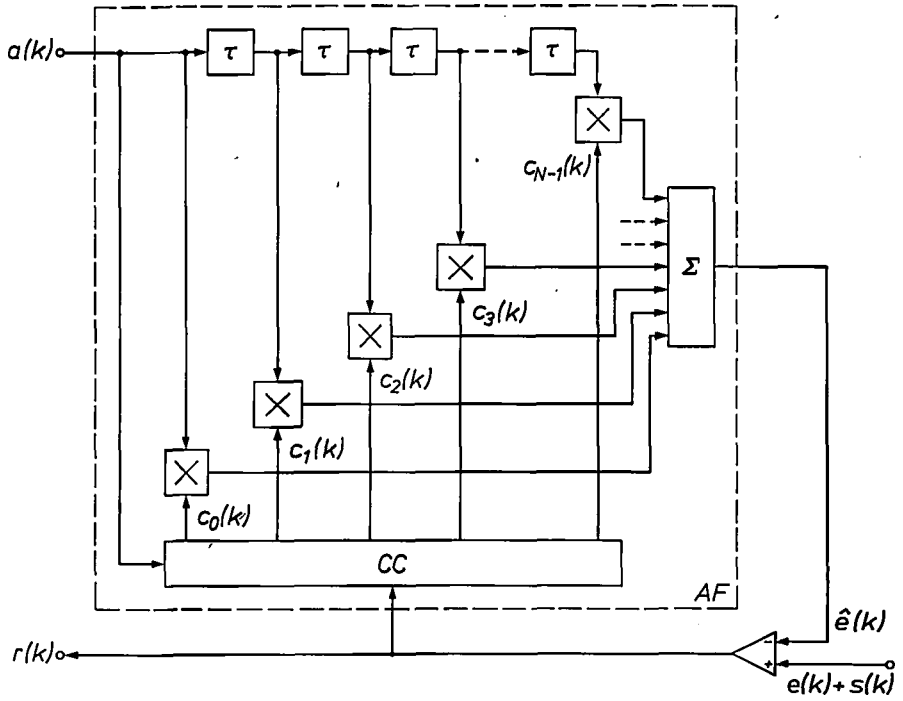


Fig. 6. Digital echo canceller consisting of an adaptive digital transversal filter *AF* and a subtracter. The circuit is derived from the one shown in fig. 5 by making the coefficients variable; their value is calculated in the circuit *CC* by means of a special algorithm. In this process use is made of the signal being filtered $a(k)$ and the output signal $r(k)$ from the subtracter.

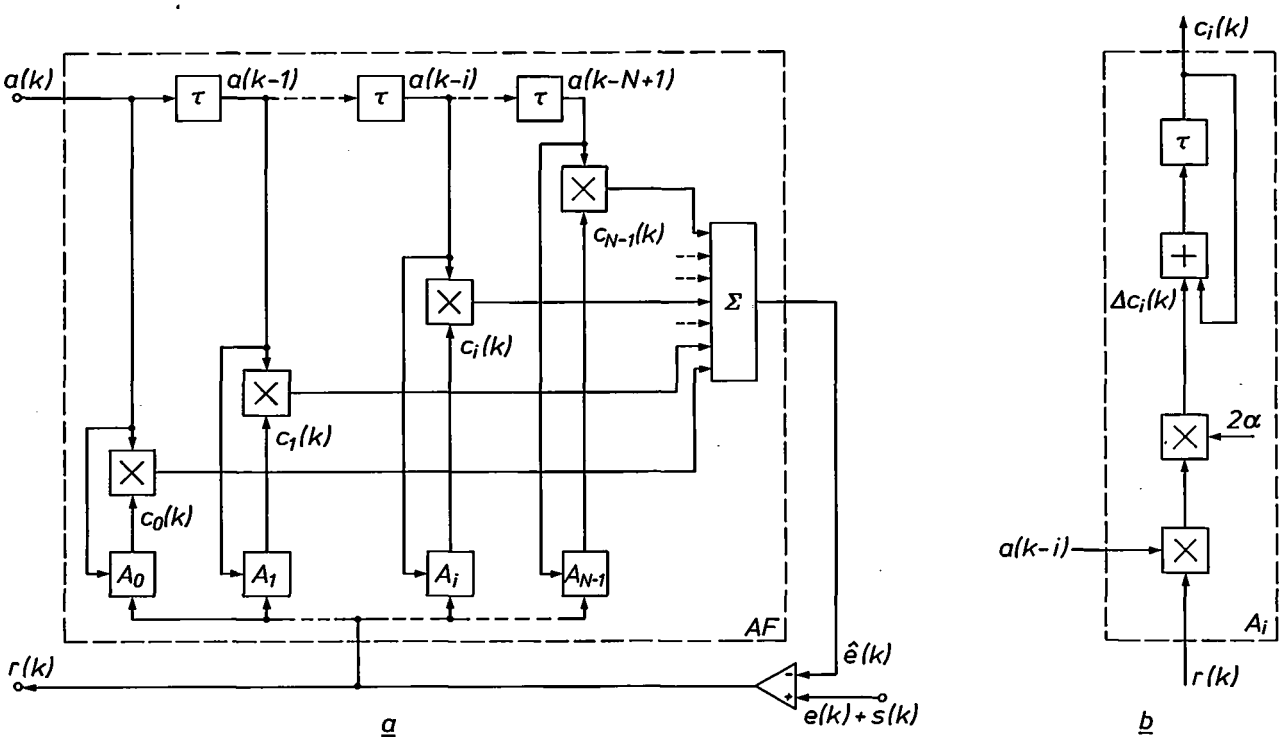


Fig. 7. *a*) A more detailed block diagram of an echo canceller based on an adaptive digital transversal filter. The value of coefficient $c_i(k)$ is calculated in the adaptation unit A_i. *b*) Block diagram of adaptation unit A_i for the stochastic-iteration algorithm, which is highly suitable for echo cancellation in a data-transmission system. α is the control constant.

Designing the filter as an integrated circuit

The adaptive digital transversal filter with N coefficients as shown in fig. 7 is not very easy to produce as a single integrated circuit, mainly because it contains a large number of multipliers, although in themselves these multipliers are simple. The sequential structure (*fig. 9*) is much more suitable; this only requires two multipliers, each operating at a rate N times faster than the multipliers in fig. 7. The numbers that have to be multiplied together in these multipliers — both the input data $a(k)$ and the coefficients $c_i(k)$ — are now stored in 'recirculation registers' (*ASR* and *CSR*) that shift N times as rapidly as the shift-register sections in fig. 7. The products supplied by the multiplier M_1 are added together in the accumulator (*Acc*) in groups of N, and form a single output sample $\hat{e}(k)$. The feedback loop of *ASR* contains a switch that allows a new data symbol to be inserted in *ASR* at regular intervals; the oldest data symbol then disappears. The feedback loop of *CSR* contains an adder that adds a correction term $2\alpha\, a(k-i)\, r(k)$ to the coefficient $c_i(k)$ during the circulation of the coefficients. These corrections are computed in multiplier M_2 from $r(k)$ and the circulating data $a(k-i)$ taken from *ASR*. Between the upper input of M_2 and the output of *ASR* there is a fixed delay D, which is necessary to compensate for the delays elsewhere in the filter (in the accumulator *Acc*, for example). The multipliers in fig. 7*b* in which $r(k)$ is multiplied by 2α are no longer found in fig. 9. This is achieved by selecting $\alpha = 2^{-m}$ (m = positive integer) so that this multiplication is carried out by shifting the samples $r(k)$ by $m-1$ bit positions. No

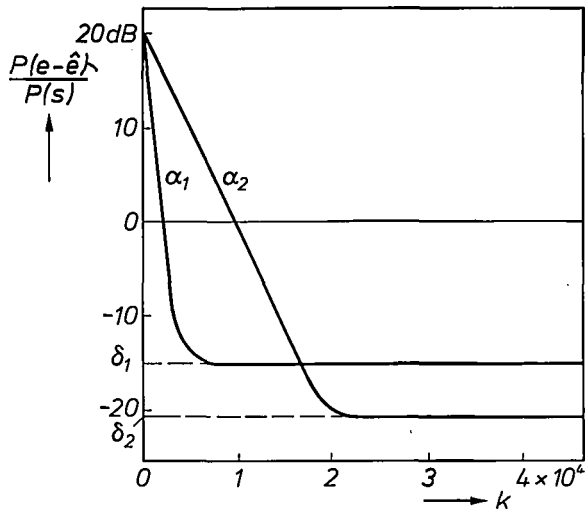


Fig. 8. The influence of the control constant α on the control behaviour of an echo canceller using the stochastic-iteration algorithm. When α is doubled, the convergence rate doubles, but the echo cancellation deteriorates by 3 dB. In the example given here $\alpha_1 = 4\alpha_2$. $P(e-\hat{e})$ is the mean power (averaged over a short period of time) of the signal $e(k)-\hat{e}(k)$. $P(s)$ is the mean power of the signal $s(k)$.

special components are then required; it is only necessary to make the appropriate connection patterns between the various digital subcircuits.

When the structure of the integrated version of the adaptive filter has been decided, it is necessary to select the magnitude of the various parameters such as the number of coefficients N, the control constant α and the number of bits (the 'word length') representing the various digital quantities in the filter. The value of N is determined by the maximum duration of the echoes that must be cancelled; the control constant α is dependent on the desired combination of final echo cancellation and convergence rate (fig. 8).

as at present. The left-hand input of M_1 and the upper input of M_2 would also have to process eight-bit words instead of the one-bit words as at present, and these multipliers would therefore be considerably more complex.

The word lengths of the residual echo $r(k)$, the coefficients $c_i(k)$ and the simulated echo $\hat{e}(k)$ have a considerable influence on the functioning of the echo canceller, but it is not immediately obvious how large they should be. We can however, draw a few general conclusions. The low value of α means that the corrections made to the coefficients at each iteration are generally small compared with the final value of

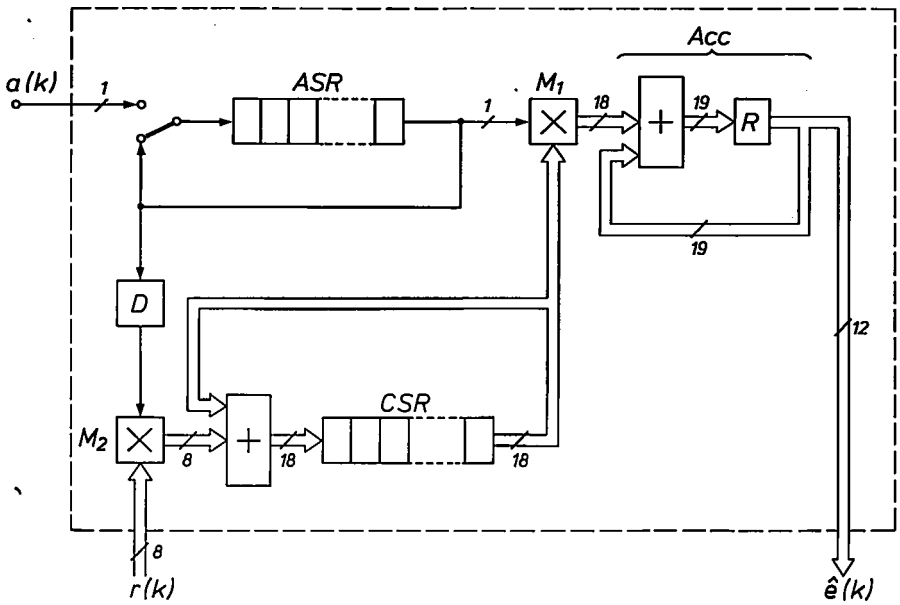


Fig. 9. Block diagram of an adaptive digital transversal filter that can be made as a single integrated circuit. This filter operates in the same way as the filter in fig. 7, but to permit integration the number of multipliers has been reduced to two (M_1 and M_2), which perform all the multiplying operations sequentially N times as fast. Both the delayed input samples $a(k-i)$ and the coefficients $c_i(k)$ are now stored in rapidly shifting recirculation registers *ASR* and *CSR*, from which they become available at exactly the right moments for further processing. The number of bits (the word length) representing the various digital quantities is indicated by the figures beside the various connections. *Acc* accumulator, *R* register, *D* delay.

We have selected $N=64$ and $\alpha=2^{-13}$; these values give satisfactory cancellation for all the echoes that occur in practice in baseband data transmission.

The word length of the signal $a(k)$ fed to the adaptive filter can be restricted to one bit if the appropriate connection is provided between transmitter and echo canceller as described earlier. Fig. 9 shows the considerable saving that this brings about compared with the situation where $a(k)$ is a digitized version of the analog output signal from the transmitter. Each sample of $a(k)$ would then consist of say eight bits, so that eight-bit words would then have to circulate in the circulation register *ASR* instead of one-bit words

the coefficients. The coefficients must therefore have a much larger word length than the corrections $2\alpha r(k)\,a(k-i)$, which themselves have the same word length as $r(k)$ because $\alpha=2^{-m}$ and $a(k-i)=1$ or -1. The final result from the accumulator *Acc* consists of the weighted sum of a large number of coefficients and will therefore have an inaccuracy equal to a multiple of the inaccuracy of each coefficient taken separately. There is little point therefore in making the word length of the simulated echo $\hat{e}(k)$ equal to the word length of the accumulator; a number of the least-significant bits of the accumulator can easily be neglected. From extensive analyses and computer

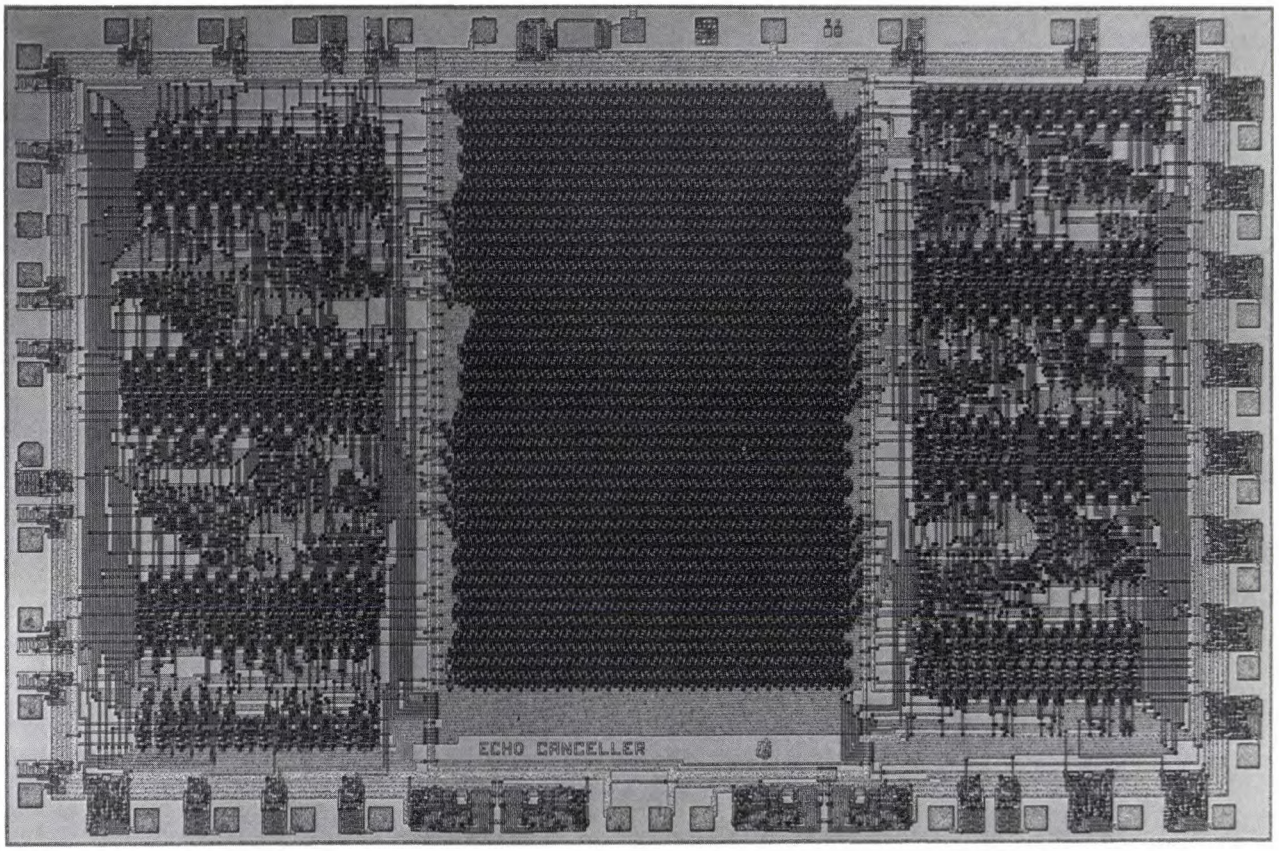


Fig. 10. Photograph of the integrated adaptive transversal filter designed for echo cancellation. The block diagram of this filter is shown in fig. 9. The regular pattern at the centre corresponds to the recirculation registers *ASR* and *CSR*. This circuit has been produced in four-phase dynamic NMOS logic and contains about 12 000 transistors on a chip area of 38 mm².

simulations [3] we have found that a good design is obtained with the following word lengths: 8 bits for the residual echo $r(k)$, 18 bits for the coefficients $c_i(k)$, 19 bits for the accumulator *Acc* and 12 bits for the simulated echo $\hat{e}(k)$ (fig. 9). The least-significant bit of $r(k)$ is equivalent to a quantization step of 40 mV, and the least-significant bit of $\hat{e}(k)$ is equivalent to a quantization step of 1.25 mV.

At first sight it may seem strange that the simulated echo $\hat{e}(k)$ is expressed in multiples of 1.25 mV, while the residual echo $r(k)$ — which ultimately determines the setting of the coefficients — is applied to the adaptive filter in multiples of not less than 40 mV. This depends on an interesting effect known as 'dither', which is due to the presence of the received signal $s(k)$ in $r(k)$; see equation (1). As a result of this effect the *instantaneous* value of the residual echo is applied to the filter with an accuracy of only 40 mV, but *on average* the corrections to the coefficients represent a much higher accuracy [3].

The adaptive filter of fig. 9 has been made as an integrated circuit in four-phase dynamic NMOS logic (*fig. 10*). It contains about 12 000 transistors (3000 logic gates) on a chip area of 38 mm². The regular pattern at the centre forms the shift registers *ASR* and *CSR*. The remainder contains the multipliers, adders and control logic. The electrical power dissipated by the IC depends on its rate of operation and is equal to about 250 mW at the maximum internal processing rate of 2.5 MHz. Depending on the transmission code used (see the following section) an echo canceller made from this IC can be used for transmission rates up to about 20 kbit/s.

Transmission codes

The way in which our integrated adaptive filter is used in an echo canceller depends to a large extent on the transmission code used. Both the connection between transmitter and adaptive filter and the interpolation factor of the adaptive filter differ from code to code. As has already been mentioned, the coupling between transmitter and echo canceller is achieved by a binary digital signal linearly related to the output

[3] N. A. M. Verhoeckx, H. C. van den Elzen, W. A. M. Snijders and P. J. van Gerwen, Digital echo cancellation for baseband data transmission, IEEE Trans. **ASSP-27**, 768-781, 1979.

signal of the transmitter. The interpolation factor of the adaptive filter is determined by the ratio of the sampling frequency of the binary digital input signal and the required sampling frequency of the output signal. This sampling frequency can be derived from the spectrum of the codes used. We shall now examine these aspects in more detail for three codes widely used for baseband data transmission. We shall therefore first describe the way in which they are generated and their spectra.

Differential biphase code

One widely used code is the differential biphase code (DBΦ); see *fig. 11a*. Starting from the 'non-return-to-zero' representation of the original data (NRZ), in which a '1' is representend by a high level and a '0' by a low level, the DBΦ code can be generated by two successive operations. The first operation is the 'differential' operation and the second the 'biphase' operation. The differential operation produces a signal *Dif*, which has a transition for every '1' in the original data sequence. A convenient feature of the *Dif* signal is that its polarity can be reversed without affecting the interpretation of the signal. The *Dif* signal can be obtained from the NRZ signal with a simple circuit containing a modulo-2 adder and a delay element. From the *Dif* signal the 'biphase' operation produces the desired DBΦ signal, which has a transition at the centre of each bit interval; for a '1' in the *Dif* signal this transition is from high to low and for a '0' the transition is from low to high. The 'biphase' operation delivers a signal that has many transitions and this is an advantage for the regeneration of the clock signal in the receiver. The circuit for executing the biphase operation consists of a sampling circuit and a linear pulse-shaping network (fig. 11*b*). The *Dif* signal is therefore linearly related to the output signal of the 'coder' and can hence be applied to the echo canceller as a binary digital signal with a sampling rate of $1/T$.

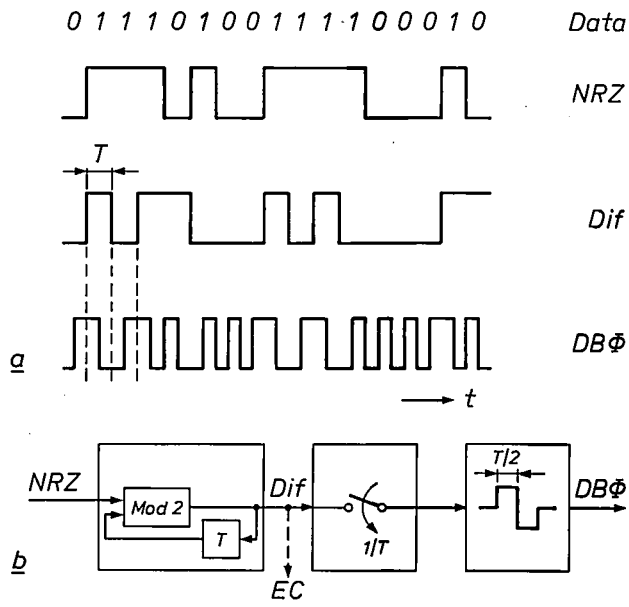


Fig. 11. *a*) The differential biphase code (DBΦ). Starting from the NRZ representation of the original data it is possible to obtain a DBΦ-coded signal by way of a differentially coded signal *Dif*. *b*) Block diagram of the circuit for DBΦ coding; only a small number of simple operations are carried out. The best way of coupling this coding circuit to an echo canceller *EC* is by means of the signal *Dif*.

Miller code

A second code that is often used is the Miller code (also called 'delay' modulation), which can be generated by cascading the biphase operation described above and a nonlinear operation with the aid of a flip-flop; see *fig. 12b*. The flip-flop delivers a signal that only has transitions at times when the biphase signal makes a transition from high to low (fig. 12*a*). In the encoded signal the distances between two transitions can only be equal to T, $\frac{3}{2}T$ or $2T$, thus providing a reasonable amount of clock information. Because of the presence of the nonlinear operation the only suitable signal that can be supplied from the transmitter to the echo canceller is the output signal from the coding circuit. However, since the transitions in this signal are not always separated by multiples of T, a sampling rate of $2/T$ must be assigned to this binary digital signal.

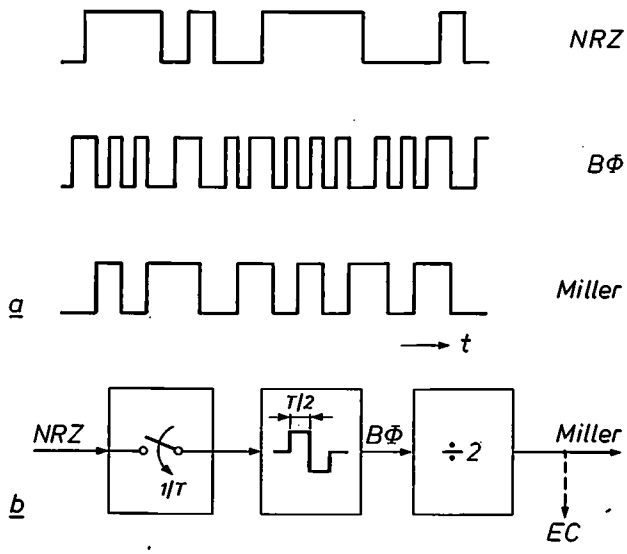


Fig. 12. *a*) The Miller code. The coded signal can be obtained from an NRZ data signal by means of a biphase operation (BΦ) and a divide-by-two operation. *b*) Circuit for Miller coding. Coupling with a possible echo canceller *EC* can only take place via the coded signal.

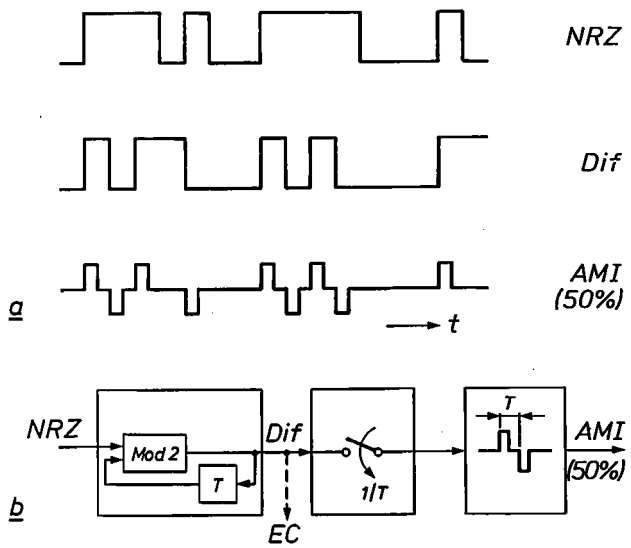


Fig. 13. *a*) The AMI(50%) code. Unlike the DBΦ code and the Miller code, this gives a three-valued coded signal. *b*) The circuit; the code can be generated in almost the same way as the DBΦ code. A comparison with fig. 11*b* shows that only the impulse response of the pulse-shaping network is different. Once again it is easiest to couple the coding circuit and the echo canceller *EC* via the *Dif* signal.

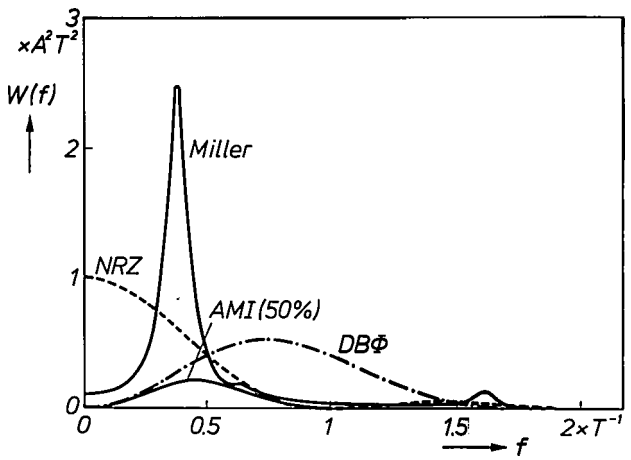


Fig. 14. Power-density spectra of data signals coded in three different ways. The bit rate of the data is always $1/T$ and all the coded signals have the same maximum positive and negative values $+A$ and $-A$. It can clearly be seen that the codes tend to suppress the low frequencies to a large extent compared with NRZ. It is also noticeable that the spectrum for DBΦ is much broader than that for Miller or AMI(50%).

AMI(50%) code

A third code that is frequently used is the 'Alternating Mark Inversion (AMI)' code, which is obtained, like the DBΦ code, by applying a second (linear) processing operation after the differential processing. In the AMI(50%) code every transition from low to high in the *Dif* signal is replaced by a positive pulse of width $T/2$ (hence the 50%) and every transition from high to low is replaced by a negative pulse (*fig. 13a*). Unlike the two codes discussed above, the AMI code gives a ternary signal. In this coded signal there is always a (negative or positive) pulse for every '1' in the original data sequence. The amount of clock information contained in the signal therefore depends on the data transmitted. To ensure that it always contains a sufficient number of 'ones', a scrambler (fig. 2) is absolutely essential for this code. A circuit for generating the AMI(50%) code is shown in fig. 13*b*. The only way in which this differs from the circuit for the DBΦ code (fig. 11*b*) is in the pulse-shaping network. It is also possible now for the *Dif* signal to be fed to the echo canceller as a binary digital signal with a sampling rate of $1/T$.

The power-density spectra of the signals obtained using the three codes discussed above are reproduced in *fig. 14*. The frequency f is plotted along the horizontal axis and the power density $W(f)$ along the vertical axis. All the signals have the same peak values $-A$ and $+A$. The spectrum of an NRZ data signal with the same peak values has been included as a reference. It can be seen that in all three codes the low frequencies are strongly attenuated or completely suppressed; this is highly desirable because the transmission path contains transformers. There is a substantial difference in the width of the spectra: in the AMI(50%) code and the Miller code most of the power is to be found in the frequency band between 0 and $1/T$ while in the DBΦ code there is a not inconsiderable contribution from the higher frequencies.

Influence of the code on the echo canceller

From the spectra in fig. 14 we can deduce that an echo canceller for the AMI(50%) code or the Miller code must be designed so that its frequency characteristic can be freely chosen in the band from 0 to $1/T$ and similarly in the band from 0 to $2/T$ for the DBΦ code. This means that the output sampling rate of the adaptive filter must be equal to $2/T$ for the AMI(50%) and the Miller codes and $4/T$ for the DBΦ code. As we have already found that the input sampling rates of the adaptive filter are $1/T$ for the AMI(50%) code, $2/T$ for the Miller code and $1/T$ for the DBΦ code, this now gives us an interpolation factor of $K=2$ for the AMI(50%) code, $K=1$ for the Miller code and $K=4$ for the DBΦ code.

Baseband modem with echo canceller

Fig. 15, shows the best way of incorporating an echo canceller in a baseband modem. The components of the original baseband modem from fig. 2 are clearly recognizable both in the transmitter *Tr* and in the receiver *Rc*. The coupling of the adaptive filter to the transmitter is dependent on the code, as explained in the previous section. In the receiver the best

location for the echo cancellation is between the input filter (LP_2) and the equalizer (Eq) so that as much noise as possible is removed from the incoming signal before echo cancellation and the equalizer is presented with an echo-free signal. The output signal of LP_2 from which the echo has to be removed is a band-limited analog signal. The signal that is supplied to the equalizer after echo cancellation must also be an analog and band-limited signal. The interfaces to the digital adaptive filter are provided by an analog-to-digital converter and a digital-to-analog converter.

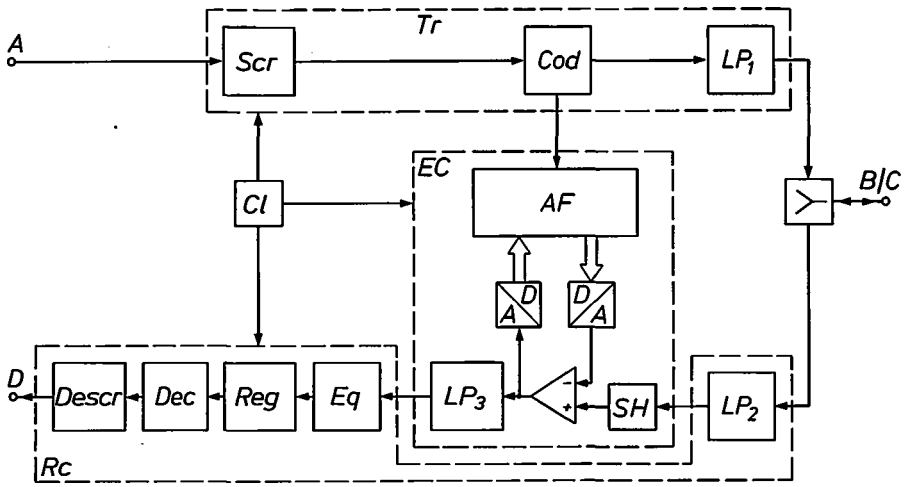


Fig. 15. Baseband modem modified for full-duplex operation on a two-wire circuit by the use of an echo canceller *EC* and a hybrid circuit. Except for these two additions, all the components are the same as those in the baseband modem for four-wire circuits (fig. 2). The adaptive filter *AF* is coupled to the analog receiver *Rc* in a special way and the actual echo cancellation (in the subtracter) is performed by means of stepped signals (see fig. 16).

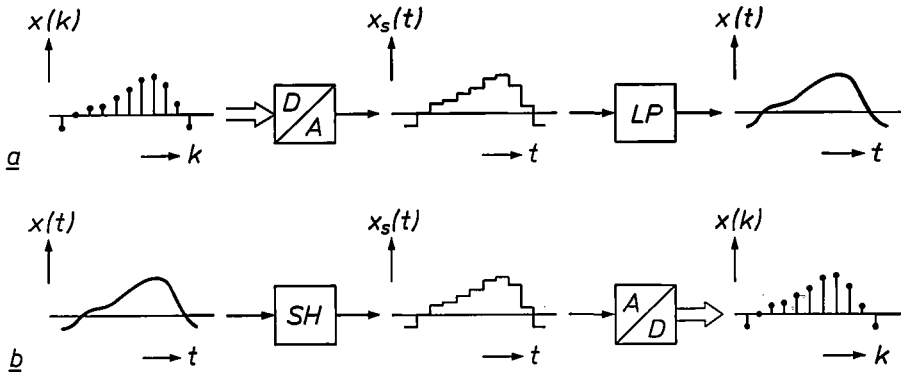


Fig. 16. The conversion of a digital signal $x(k)$ into a band-limited analog signal $x(t)$ and vice versa normally takes place in two stages. The intermediate result is a stepped signal $x_s(t)$, which differs from $x(t)$ because of the presence of spectral components at higher frequencies. *a*) The circuit, known in practice as a digital-to-analog converter, converts the signal $x(k)$ into a signal $x_s(t)$; from this signal it is then possible to obtain the band-limited analog signal $x(t)$ by means of a lowpass filter *LP*. *b*) The digitization of an analog signal $x(t)$ starts with a sample-and-hold operation *SH* that produces a stepped signal $x_s(t)$. This signal is then processed in an analog-to-digital converter to form the samples of the digital signal $x(k)$.

The conversion of a digital signal into a band-limited analog signal requires the use of a lowpass filter. Because it is undesirable to have this filter in the control loop of the echo canceller, this operation is now performed by the filter LP_3, which is located outside the actual control loop. The output signal of the digital-to-analog converter forms a stepped approximation to the desired analog signal (*fig. 16*). By using a sample-and-hold circuit *SH* to give the output signal of filter LP_2 a similar stepped shape, the actual echo cancellation is then carried out by taking the dif-

ference between two similar signals. It can be shown [3] that this simplifies the requirements for filter LP_3, which is used in producing the desired echo-free band-limited analog signal.

The photographs in *fig. 17* illustrate the performance of a baseband modem with an echo canceller in operation for two-wire full-duplex traffic. The photographs show 'eye patterns'. These are obtained by reproducing a data signal on an oscilloscope with the time-base interval made equal to one data-symbol interval. Images of successive segments of the data signal are then superimposed to form an image that looks like one or more eyes (hence the name 'eye pattern'). If the eyes are 'open', good regeneration of the data signal is possible; as the eyes 'close', the quality of the data signal deteriorates.

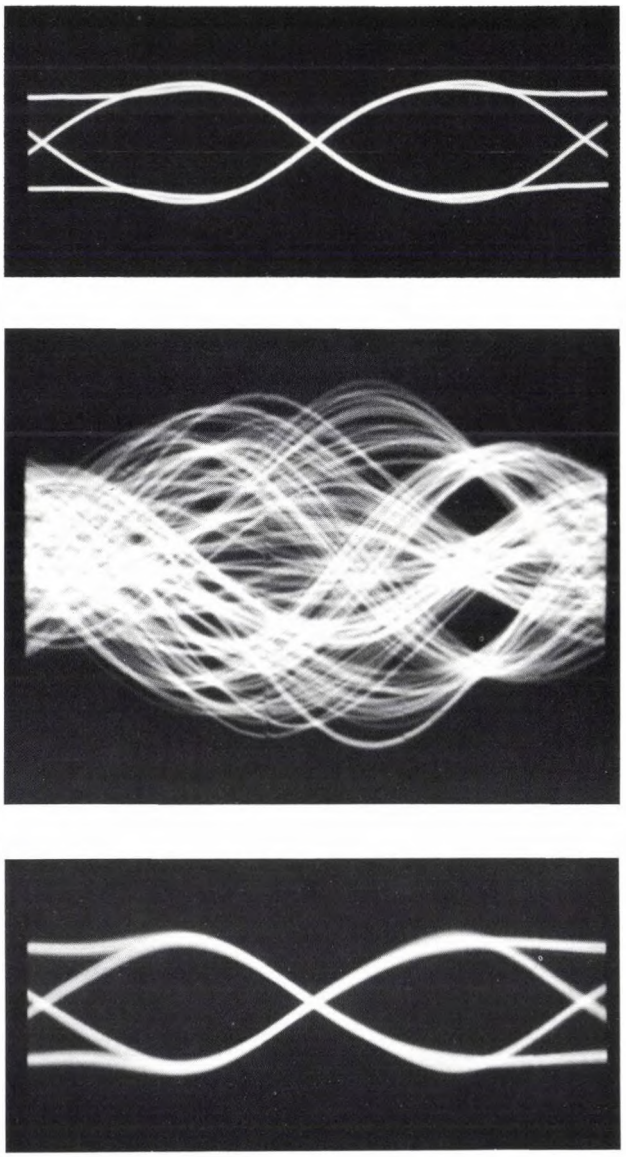


Fig. 17. Eye patterns of signals at the input of the regenerator *Reg* in the baseband modem of fig. 15. The echo canceller here contains the integrated adaptive digital filter of fig. 10. *a*) Ideal case; there are no echoes. The eyes of the eye pattern are 'open' and high-quality regeneration can take place. *b*) Practical situation in which the desired signal is considerably perturbed by echoes, but there is no echo cancellation. The eye pattern is so distorted that good regeneration can no longer be achieved. Under these circumstances full-duplex data traffic is impossible on the two-wire circuit. *c*) The same situation as in (*b*), but now with the echo canceller in operation. The eye pattern is almost as good as in (*a*) and high-quality full-duplex data communication is possible on the two-wire circuit.

All the eye patterns shown in fig. 17 have been recorded at the output of the equalizer *Eq* in fig. 15 and relate to a DBΦ-coded data signal. If there are no echoes, an open eye pattern is obtained, like the one in fig. 17*a*. Fig. 17*b* shows the output signal of the equalizer for the case where there are echoes but echo cancellation is not used. The echoes here are more than twice as large as the desired signal. The eye pattern is perturbed to such an extent that good regeneration is impossible. Finally, fig. 17*c* shows the eye pattern obtained under the same conditions as in fig. 17*b*, but this time after echo cancellation with an integrated adaptive digital filter as described here. The echoes have been reduced to a negligible level, 10 to 20 times less than the desired signal. Good regeneration is now possible even though there are echoes at the input to the receiver.

It is interesting to consider the minimum r.m.s. values of the signals that can occur at the output of the equalizer. For the desired signals this minimum value is 10 mV and there is a residual echo of only 0.5 to 1 mV. However, the quantization step of the analog-to-digital converter in fig. 15 in which the residual echo is digitized is much larger, 40 mV. Echo cancellation to such a high accuracy can still be obtained with such rough quantization because of the dither effect encountered earlier.

Future prospects

The integrated adaptive digital filter that we have made has been designed for use in echo cancellers for baseband modems with a maximum transmission rate of 20 kbit/s. However, even at the present state of the technology it is already possible to construct this kind of circuit for much higher bit rates (e.g. 200 kbit/s). Existing two-wire circuits (without loading coils), on the other hand, can in principle be used for data transmission at these higher rates provided that the length of the circuit is restricted to a few kilometres. The combination of these two as yet unused features may lead to new applications to satisfy the requirements that will inevitably develop. One of these is the combined transmission of digitized speech and all kinds of other digital information (data, telex, Viewdata, electronic mail, etc.) [4]. This can be achieved

[4] W. A. Kaiser and H. T. Hagmeyer, Digital two-wire local connection providing office subscribers with speech, data and new teleinformation services, Conf. Rec. IEEE Int. Symp. on Subscriber loops and services, Atlanta 1978, pp. 126-130.

by transmitting various digital signals in time-division multiplex. Each of the signal sources is then periodically permitted to make use of the transmission path for a short time. The transmitted signal consists of 'frames' of a given number of bits, and hence of fixed duration, with certain parts of the frame reserved for each of the signal sources. We shall give an actual example of such a frame structure in a moment.

A digital coding method that is commonly used for speech (pulse-code modulation or PCM) provides a bit rate of 64 kbit/s. With the bits required for signalling and synchronization a fully digital telephone signal of say 96 kbit/s is obtained.

An attractive new system for the combined signal transmission that we have mentioned above is obtained by using 192 kbit/s for each of the two transmission directions and by applying echo cancellation to provide full-duplex operation. This system provides 64 kbit/s for speech transmission in both directions and an additional 64 kbit/s for the various other types of digital information. If desired, speech alone can be transmitted; two telephone conversations can then be made simultaneously on a single 2-wire circuit.

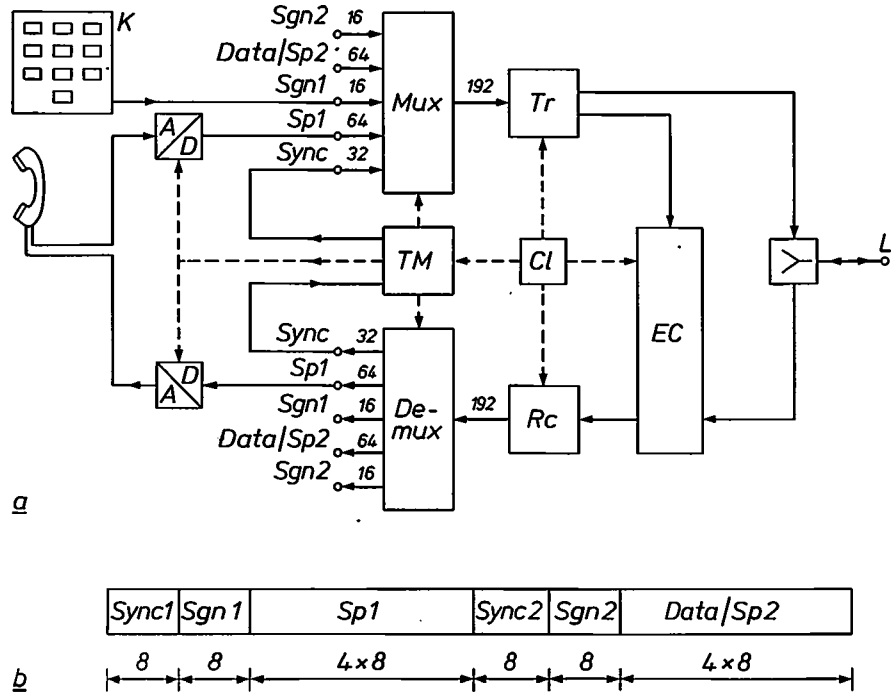


Fig. 18. Another application of echo cancellation. Because the bandwidth is used very efficiently with echo cancellation, full-duplex digital communication can be achieved at a rate of 192 kbit/s on two-wire circuits. This high information capacity can be used for transmitting a number of signals in time-division multiplex. Correct selection of bit rates and frame structure gives a very versatile transmission system. *a*) Block diagram of the subscriber's equipment. The right-hand part of the circuit performs the same functions (*Tr, Rc, EC, Cl* and hybrid) as the modem in fig. 15. An important part of the subscriber's equipment is now formed by the multiplex and demultiplex circuits *Mux* and *Demux* with their associated control and timing circuit *TM*. In the example given here the multiplex signal can consist of two 64 kbit/s information signals (*Sp1* and *Data/ Sp2*), two 16 kbit/s signalling signals (*Sgn1* and *Sgn2*) and one 32 kbit/s synchronization signal (*Sync*). The drawing indicates symbolically how digitized speech can be transmitted by means of signal *Sp1* and analog-to-digital and digital-to-analog converters, while the signal *Sgn1* can be obtained from a keyboard *K*. The figures by the connections represent bit rates in kbit/s. *b*) Frame structure of the 192 kbit/s multiplex signal. Each frame contains 96 bits and has two identical halves, in which 8 synchronization bits and 8 signalling bits are followed by 4×8 bits for the transmission of information (speech or data).

Fig. 18 shows the block diagram of the subscriber's equipment in this system. The right-hand half of the figure represents the actual transmitter/receiver unit consisting of transmitter, receiver, echo canceller, hybrid circuit and clock circuits. The left-hand side of the figure is a schematic illustration of the multiplex/demultiplex operation. In the multiplexer, two digital information streams of 64 kbit/s (speech or data) are combined with two signalling streams of 16 kbit/s and one synchronization signal of 32 kbit/s to form a

single multiplex signal of 192 kbit/s (the 64 kbit/s data signal may already be a multiplex signal consisting say of a number of telex signals; equipment for producing such multiplex signals is already in use in existing data networks). A possible frame structure for the 192 kbit/s multiplex signal is given in fig. 18*b*. Each frame consists of two almost identical halves, each starting with 8 bits of synchronization information. This is followed by 8 bits for signalling and 4×8 bits for speech or data. The synchronization information must include an indication for differentiating between the two halves of the frame. This is done by making the first seven synchronization bits in the first half of the frame the logical inversion of the first seven synchronization bits in the second half of the frame. The eighth bit in each half of the frame is used to indicate whether speech or data is to follow.

In the demultiplexer at the receiving end the operation is the converse of that in the multiplexer: the incoming 192 kbit/s signal is split up again into its various original components. A very important function is performed here by the timing circuit (*TM*), which ensures that all the operations are performed at exactly the right moments; in the receiver unit, the necessary information for this must be derived from the incoming signal.

The potential application that we have been describing in this section clearly shows how echo cancellation can be used to make full-duplex operation a practical proposition for digital transmission at rates as high as 200 kbit/s on two-wire circuits in a local telephone network. This is because echo cancellation permits a more efficient use of the available bandwidth than is possible with other known methods.

Summary. The increasing volume of data traffic is leading to the introduction of special data networks; these are better able to satisfy the specific requirements of data transmission than the conventional telephone networks that have so far been used. For the connection between a subscriber and the central part of the data network use is made of the existing telephone subscriber lines and 'baseband' modems. By using echo cancellers it is possible to transmit data signals on a single subscriber line (i.e. one pair of copper wires) simultaneously in both directions and in the same frequency band. Such an echo canceller can be built from an adaptive filter designed as an integrated circuit. This IC, which can be used up to a maximum transmission rate of 20 kbit/s, contains about 12 000 transistors on a chip area of about 38 mm². The use of echo cancellers on an even larger scale in the 'digital telephone' is also envisaged for the not too distant future. Because the available bandwidth is used efficiently simultaneous two-way transmission of 64 kbit/s for speech plus say 64 kbit/s for other digital signals is possible on the local telephone network.

Scientific publications

These publications are contributed by staff of laboratories and plants which form part of or cooperate with enterprises of the Philips group of companies, particularly by staff of the following research laboratories:

Philips Research Laboratories, Eindhoven, The Netherlands	*E*
Philips Research Laboratories, Redhill, Surrey, England	*R*
Laboratoires d'Electronique et de Physique Appliquée, 3 avenue Descartes, 94450 Limeil-Brévannes, France	*L*
Philips GmbH Forschungslaboratorium Aachen, Weißhausstraße, 51 Aachen, Germany	*A*
Philips GmbH Forschungslaboratorium Hamburg, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany	*H*
Philips Research Laboratory Brussels, 2 avenue Van Becelaere, 1170 Brussels (Boitsfort), Belgium	*B*
Philips Laboratories, N.A.P.C., 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A.	*N*

P. M. Asbeck, D. A. Cammack & J. J. Daniele: Non-Gaussian fundamental mode patterns in narrow-stripe-geometry lasers.
Appl. Phys. Letters **33**, 504-506, 1978 (No. 6). *N*

M. Audier, J. C. Delmotte & J. P. Boutot: Multiplicateur à galette de microcanaux: amélioration des performances de gain et de dynamique de détection.
Rev. Phys. appl. **13**, 188-194, 1978 (No. 4). *L*

M. Auphan, R. Goudin & G. Dale (CHU de Strasbourg, Service Central de Radiologie): Amélioration du diagnostic échographique grâce à la focalisation en poursuite d'échos.
C. R . BIOSIGMA 78, Coll. Int. sur les Signaux et les images en médecine et en biologie, Paris 1978, tome 2, pp. 122-127. *L*

R. N. Bates: Microwave low noise mixers with image and sum frequency enhancement.
Conf. on Radio receivers and associated systems, Southampton 1978 (IERE Conf. Proc. No. 40), pp. 317-322. *R*

R. N. Bates & R. E. Pearson: Designing bandstop filters for microwave frequencies.
Electronic Engng. **50**, April 1978, 39-41 (No. 604). *R*

N. M. Beekmans, M. Brouha & H. J. van Daal: The effect of high pressure on the resistivity of Cr-Al disordered alloys.
Transition metals 1977, Int. Conf. Toronto (Inst. Phys. Conf. Ser. No. 39), pp. 429-432; 1978. *E*

V. Belevitch & J. Boersma (Eindhoven University of Technology): The Bessel ratio $K_{\nu+1}(z)/K_{\nu}(z)$ as a passive impedance.
Philips J. Res. **34**, 163-173, 1979 (No. 3/4). *B*

C. Belouet, J. Hervo, R. Martres, Ngo Tich Phuoc & M. Pertus: Growth and characteristics of polysilicon layers achieved by the ribbon-against-drop (RAD) pulling process.
13th IEEE Photovoltaic Specialists Conf., Washington D.C. 1978, pp. 131-136. *L*

J. Bloem: Silicon growth mechanisms.
Acta Electronica **21**, 201-208, 1978 (No. 3). *E*

P. Blood & J. W. Orton: The electrical characterisation of semiconductors.
Rep. Prog. Phys. **41**, 157-257, 1978 (No. 2). *R*

B. Bölger: High-resolution Raman spectroscopy with a tunable laser.
Laser spectroscopy, Lecture Notes in Physics **43**, 460, 1975. *E*

B. Bölger: Moderne spectroscopische methoden met lasers.
Chem. Weekbl. Mag. 1977, pp. m 55, 57, 59, 60 (Feb.). *E*

B. Bölger, L. Baede & H. M. Gibbs (Bell Laboratories, Murray Hill, N.J.): Production of tunable 300 watt, nanosecond, transform limited optical pulses and their application to coherent pulse propagation.
Optics Comm. **18**, 67-68, 1976 (No. 1). *E*

L. J. M. Bollen: Epitaxial silicon, state of the art.
Acta Electronica **21**, 185-199, 1978 (No. 3). *E*

J. van den Boomgaard: Eutectic-like and peritectic-like reactions in ternary systems at a constant pressure, Part II. Systems containing one liquid phase and three solid phases but no compounds of the components.
Philips J. Res. **34**, 48-77, 1979 (No. 1/2). *E*

B. C. Bouma & A. C. Roelofs: An experimental determination of the forward-biased emitter-base capacitance.
Solid-State Electronics **21**, 833-836, 1978 (No. 6). *E*

P. W. J. M. Boumans: 'Plasma sources' for multi-element analysis of solutions: capabilities, limitations, and perspectives.
Mikrochim. Acta 1978 I, 399-412 (No. 3/4). *E*

P. W. J. M. Boumans, L. C. Bastings, F. J. de Boer & L. W. J. van Kollenburg: ICP-atomic emission spectrometry as a tool for flexible single-element analysis of non-routine samples.
Fresenius' Z. anal. Chemie **291**, 10-19, 1978 (No. 1). *E*

D. J. Breed, A. M. J. van der Heijden, H. Logmans & A. B. Voermans: High domain-wall velocities in $(Gd,Lu)_3(Fe,Mn,Al)_5O_{12}$ garnet epilayers with orthorhombic anisotropy.
J. appl. Phys. **49**, 939-941, 1978 (No. 2). *E*

C. H. J. van den Brekel: Characterization and morphology of chemical vapour deposition of silicon.
Acta Electronica **21**, 209-220, 1978 (No. 3). *E*

C. H. J. van den Brekel & A. K. Jansen: Morphological stability analysis in chemical vapour deposition processes, I, II.
J. Crystal Growth **43**, 364-370, 371-377, 1978 (No. 3). *E*

C. H. J. van den Brekel & A. K. Jansen: Interface morphology in chemical vapour deposition on profiled substrates.
J. Crystal Growth **43**, 488-496, 1978 (No. 4). *E*

J. W. Broer: Integrating text and figures.
Proc. 25th Int. tech. Comm. Conf., Dallas 1978, pp. 36-40. *E*

A. Broese van Groenou: Scratch-modified magnetic domain patterns in garnet films.
Appl. Phys. **15**, 385-392, 1978 (No. 4). *E*

H. H. Brongersma & T. M. Buck (Bell Laboratories, Murray Hill, N.J.): Low-energy ion scattering (LEIS) for composition and structure analysis of the outer surface.
Nucl. Instr. Meth. **149**, 569-575, 1978 (Nos. 1-3). *E*

M. R. Brozel*, J. B. Clegg & R. C. Newman* (* J. J. Thomson Physical Laboratory, Reading): Carbon, oxygen and silicon impurities in gallium arsenide.
J. Physics D **11**, 1331-1339, 1978 (No. 9). *R*

R. K. Bullough, P. J. Caudrey, J. D. Gibbon, S. Duckworth (all with University of Manchester Institute of Science and Technology), **H. M. Gibbs** (Bell Laboratories, Murray Hill, N.J.), **B. Bölger & L. Baede:** Wobbling and leap frogging in self-induced transparency.
Optics Comm. **18**, 200-201, 1976 (No. 2). *E*

A. L. J. Burgmans, B. Bölger & M. F. H. Schuurmans: Anomaly in the fluorescence excitation spectrum of an atomic vapour near an interface.
Proc. 3rd Nat. Quantum Electronics Conf., Southampton 1977, paper 8, p. 20. *E*

A. L. J. Burgmans & M. F. H. Schuurmans: Spectral narrowing in selective reflection.
Optics Comm. **18**, 137-138, 1976 (No. 1). *E*

K. H. J. Buschow & A. M. van der Kraan (Interuniversitair Reactor Instituut, Delft): Magnetic ordering in ternary rare earth iron aluminium compounds (RFe_4Al_8).
J. Physics F **8**, 921-932, 1978 (No. 5). *E*

K. H. J. Buschow & R. C. Sherwood (Bell Laboratories, Murray Hill, N.J.): Effect of H_2 absorption on the magnetic properties of rare-earth transition metal compounds.
J. appl. Phys. **49**, 1480-1485, 1978 (No. 3, Part II). *E*

K. H. J. Buschow, R. C. Sherwood*, F. S. L. Hsu* & K. Knorr** (* Bell Laboratories, Murray Hill, N.J.; ** University of Mainz): Magnetic properties of rare-earth magnesium compounds of the type RMg_2.
J. appl. Phys. **49**, 1510-1512, 1978 (No. 3, Part II). *E*

K. H. J. Buschow, J. H. Wernick* & G. Y. Chin* (* Bell Laboratories, Murray Hill, N.J.): Note on the Hf-Co phase diagram.
J. less-common Met. **59**, 61-67, 1978 (No. 1). *E*

K. H. J. Buschow, G. Will* & M. O. Bargouth* (*Universität Bonn): Magnetic properties of $GdMg_3$ and $TbMg_3$.
J. Physics C **11**, 2405-2413, 1978 (No. 12). *E*

M. C. W. van Buul: Hybrid D-PCM, a combination of PCM and DPCM.
IEEE Trans. **COM-26**, 362-368, 1978 (No. 3). *E*

P. J. Cameron (Merton College, Oxford), **P. Delsarte & J.-M. Goethals:** Hemisystems, orthogonal configurations, and dissipative conference matrices.
Philips J. Res. **34**, 147-162, 1979 (No. 3/4). *B*

K. Carl & K. H. Härdtl: Electrical after-effects in $Pb(Ti,Zr)O_3$ ceramics.
Ferroelectrics **17**, 473-486, 1978 (No. 3/4). *A*

F. M. A. Carpay: Normal grain growth.
Ber. Bunsen-Ges. phys. Chemie **82**, 306-308, 1978 (No. 3). *E*

W. A. Cense, M. Klerk & W. Albers: Aligned *in situ* growth of the fibrous Ni-Nb_2O_5 composite by preferential internal oxidation of amorphous Nb_2Ni_3.
J. Mat. Sci. **12**, 2184-2188, 1977 (No. 11). *E*

C. Clausen & U. Killat: Holographically stored x-ray images: gray-tone reproduction.
Med. Phys. **5**, 181-187, 1978 (No. 3). *H*

D. J. Coe: Changes in effective channel length due to hot-electron trapping in short-channel m.o.s.t.s.
IEE J. Solid-St. Electron Dev. **2**, 57-61, 1978 (No. 2). *R*

R. L. Cohen, K. W. West (both with Bell Laboratories, Murray Hill, N.J.) **& K. H. J. Buschow:** Degradation of hydrogen absorbing rare earth intermetallics by cycling.
Solid State Comm. **25**, 293-295, 1978 (No. 5). *E*

R. W. Cooper & D. H. Paxman: Measurement of charge-carrier behaviour in PIN diodes using a laser technique.
Solid-State Electronics **21**, 865-869, 1978 (No. 6). *R*

H. Dammann: Color separation gratings.
Appl. Optics **17**, 2273-2279, 1978 (No. 15). *H*

M. Davio & G. Bioul: Efficiency of pipelined combinational circuits.
Digital Processes **4**, 3-16, 1978 (No. 1). *B*

M. Davio & J.-P. Deschamps: Addition in signed digit number systems.
Proc. 8th Int. Symp. on Multiple-valued logic, Rosemont, Ill., 1978, pp. 104-113. *B*

M. Davio, J.-P. Deschamps & A. Thayse: Discrete and switching functions.
Book, xx + 729 pp.; McGraw-Hill, New York 1978. *B*

M. Davio & A. Thayse: Algorithms for minimal-length schedules.
Philips J. Res. **34**, 26-47, 1979 (No. 1/2). *B*

E. H. L. J. Dekker, K. L. L. van Mierloo & R. de Werdt: Conductor and transfer-gate performance of single-mask field-current access magnetic-bubble devices.
J. appl. Phys. **49**, 1927-1929, 1978 (No. 3, Part II). *E*

N. H. Dekkers & H. de Lang: Comment on 'Scanning transmission electron microscopy of thin specimens' by J. M. Cowley.
Ultramicroscopy **3**, 101-102, 1978 (No. 1). *E*

P. Delsarte, Y. Genin & Y. Kamp: Asymptotic properties of the parametrization of block-Toeplitz systems.
1978 IEEE Int. Symp. on Circuits and systems, New York, pp. 968-971. *B*

J. G. Dil, P. F. Greve & W. Mesman: Measurement of steep aspheric surfaces.
Appl. Optics **17**, 553-557, 1978 (No. 4). *E*

A. G. Dirks & H. J. Leamy: Microstructure and magnetism in amorphous rare earth - transition metal thin films.
J. appl. Phys. **49**, 1735-1737, 1978 (No. 3, Part II). *E*

R. van den Doel, M. F. H. Schuurmans & K. Weiss: Playing with basic concepts of physics. The research of D. Polder.
Philips J. Res. **34**, 91-106, 1979 (No. 3/4). *E*

A. Dumont: Application des technologies mixtes analogiques-numériques à la profilométrie Doppler.
C.R. BIOSIGMA 78, Coll. Int. sur les Signaux et les images en médecine et en biologie, Paris 1978, tome 1, pp. 418-420. *L*

H. Durand: Perspectives et limitations économiques de l'emploi de la conversion photovoltaïque.
Bull. A.I.S.M./I.A.L.A. Bull. 1978-2, pp. 15-21. *L*

B. C. Easton: Chemical vapour deposition of indium phosphide.
Acta Electronica **21**, 151-158, 1978 (No. 2). *R*

L. D. J. Eggermont & E. C. Dijkmans: Signal/quantizing-distortion ratio measurements of fast-adaptive delta modulation systems.
1978 IEEE Int. Conf. on Acoustics, speech & signal processing, Tulsa, Okla., pp. 569-572. *E*

W. van Erk: The growth kinetics of garnet liquid phase epitaxy using horizontal dipping.
J. Crystal Growth **43**, 446-456, 1978 (No. 4). *E*

L. J. M. Esser & L. G. M. Heldens: Intermitterend ladingstransport in de halfgeleider en zijn toepassingen.
Polytechn. T. Elektr. **33**, 2-16, 145-153, 209-219, 1978 (Nos. 1, 3, 4). *E*

E. Fabre, Y. Baudet & S. Makkram Ebeid: Solar cells on R.A.D. polycrystalline silicon.
13th IEEE Photovoltaic Specialists Conf., Washington D.C. 1978, pp. 1101-1105. *L*

C. T. Foxon: Molecular beam epitaxy.
Acta Electronica **21**, 139-150, 1978 (No. 2). *R*

G. Frank, E. Kauer & H. Köstlin: Transparente wärmereflektierende Schichten auf der Basis hochdotierter Halbleiter.
Phys. Blätter **34**, 106-115, 1978 (No. 3). *A*

P. E. C. Franken & H. van Doveren: Determination of the grain boundary composition of soft ferrites by Auger electron spectroscopy.
Ber. Dtsch. Keram. Ges. **55**, 287-289, 1978 (No. 6). *E*

N. V. Franssen: On the acoustical transmission function in rooms.
Acustica **40**, 15-20, 1978 (No. 1). *E*

P. J. Garner: Voltage controlled crystal oscillators.
Conf. on Radio receivers and associated systems, Southampton 1978 (IERE Conf. Proc. No. 40), pp. 227-235. *R*

P. J. Garner & R. F. Mitchell: A rapid break-in system for simplex radio channels.
Conf. on Radio receivers and associated systems, Southampton 1978 (IERE Conf. Proc. No. 40), pp. 11-18. *R*

J. A. Geurst: Vorticity in two-fluid hydrodynamics.
Symp. on Applied Mathematics dedicated to the late Prof. Dr. R. Timman, Delft 1978, pp. 91-97. *E*

H. M. Gibbs (Bell Laboratories, Murray Hill, N.J.), **B. Bölger & L. Baede:** On-resonance self-focusing of optical pulses propagating coherently in sodium.
Optics Comm. **18**, 199-200, 1976 (No. 2). *E*

H. M. Gibbs (Bell Laboratories, Murray Hill, N.J.), **B. Bölger, F. P. Mattar*, M. C. Newstein*, G. Forster** & P. E. Toschek**** (* Polytechnic Institute of New York, Farmingdale, N.Y.; ** University of Heidelberg): Coherent on-resonance self-focusing of optical pulses in absorbers.
Phys. Rev. Letters **37**, 1743-1746, 1976 (No. 26). *E*

J. T. H. Goossen & J. G. Kloosterboer: Photolysis and hydrolysis of adenosine 5'-phosphates.
Photochem. Photobiol. **27**, 703-708, 1978 (No. 6). *E*

M. Goscianski: Hybrid tilt alignment for liquid crystal matrix displays.
Philips J. Res. **34**, 78-87, 1979 (No. 1/2). *L*

P. Groenveld & A. de Jong (Dr Neher Laboratory, PTT, Leidschendam): A simple r.f. immunity test set-up.
Electromagnetic compatibility 1977, Proc. 2nd Symp. and tech. Exhib., Montreux, pp. 233-239. *E*

P. C. M. Gubbens, A. M. van der Kraan (both with Interuniversitair Reactor Instituut, Delft) **& K. H. J. Buschow:** Concentration dependence of the Fe moments in rare earth iron compounds.
Solid State Comm. **26**, 107-110, 1978 (No. 2). *E*

G. J. van Gurp: The growth of metal silicide layers on silicon.
Semiconductor silicon 1977, Electrochem. Soc. Proc. Vol. 77-2, pp. 342-358. *E*

W. van Haeringen, F. A. Staas & J. A. Geurst: On ^{3}He-^{4}He dilution refrigerators, Parts I and II.
Philips J. Res. **34**, 107-142, 1979 (No. 3/4). *E*

H.-J. Hagemann: Electrical properties of acceptor-doped $BaTiO_3$ ceramic.
Ber. Dtsch. Keram. Ges. **55**, 353-355, 1978 (No. 7). *A*

J. A. G. Halé & P. Saraga: Digital image processing. Pattern recognition: ideas in practice, ed. B. G. Batchelor, pp. 177-202; Plenum Press, New York 1978. *R*

J.-P. Hallais: Croissance épitaxiale de semiconducteurs III-V à partir de composés organométalliques et d'hydrures.
Acta Electronica **21**, 129-138, 1978 (No. 2). *L*

G. Harding: Spurious resolution and transfer function *(comment)*.
Phys. Med. Biol. **23**, 526-528, 1978 (No. 3). *H*

G. Harding, U. Bertram & H. Weiss: Towards optimum blurring in spiral tomography.
Med. Phys. **5**, 280-284, 1978 (No. 4). *H*

P. Harrop & T. A. C. M. Claasen: Modelling of an f.e.t. mixer.
Electronics Letters **14**, 369-370, 1978 (No. 12). *L, E*

J. C. M. Henning & J. H. den Boef: Reply to 'Comments on Strain-modulated electron spin resonance of Co^{2+} in MgO: A comparison of extensional and flexural modes'.
Phys. Rev. B **17**, 3019, 1978 (No. 7). *E*

D. Hennings: Liquid phase sintering of barium titanate.
Ber. Dtsch. Keram. Ges. **55**, 359-360, 1978 (No. 7). *A*

A. van Herk: Side-fringing response of magnetic reproducing heads.
J. Audio Engng. Soc. **26**, 209-211, 1978 (No. 4). *E*

H. Heyns & J. G. van Santen: The resistive gate CTD area-image sensor.
IEEE Trans. **ED-25**, 135-139, 1978 (No. 2). *E*

L. Hollan: La croissance épitaxiale de GaAs en phase vapeur.
Acta Electronica **21**, 117-127, 1978 (No. 2). *L*

L. Honds & K. H. Meyer: 2-poliger Spaltpol-Synchronmotor mit Hystereseläufer.
Feinwerktechnik & Messtechnik **86**, 168-171, 1978 (No. 4). *A*

L. Honds & K. H. Meyer: Beeinflussung des Drehmoments von Hysteresemotoren durch Oberwellen.
Feinwerktechnik & Messtechnik **86**, 172-176, 1978 (No. 4). *A*

J. B. Hughes: Comments on 'Improved feedback ECL gate with low delay-power product for the subnanosecond region'.
IEEE J. **SC-13**, 276-278, 1978 (No. 2). *R*

A. P. Hulst & C. Lasance (Philips Elcoma Division, Eindhoven): Ultrasonic bonding of insulated wire.
Welding J. **57**, Feb. 1978, 19-25 (No. 2). *E*

C. Hurtes, M. Boulou, A. Mitonneau & D. Bois: Deep-level spectroscopy in high-resistivity materials.
Appl. Phys. Letters **32**, 821-823, 1978 (No. 12). *L*

H. Ihrig: The PTC effect of semiconducting $BaTiO_3$ ceramics as a function of the titanium excess.
Ber. Dtsch. Keram. Ges. **55**, 319-321, 1978 (No. 6). *A*

H. Ihrig & D. Hennings: Electrical transport properties of *n*-type $BaTiO_3$.
Phys. Rev. B **17**, 4593-4599, 1978 (No. 12). *A*

G. Jacob: Croissance épitaxiale de GaN.
Acta Electronica **21**, 159-165, 1978 (No. 2). *L*

W. P. A. Joosen: Fluid mechanical aspects in vapour deposition processes.
Symp. on Applied Mathematics dedicated to the late Prof. Dr. R. Timman, Delft 1978, pp. 69-77. *E*

J. T. C. van Kemenade & R. K. Eynthoven: Direct determination of barrier voltage in ZnO varistors.
Ber. Dtsch. Keram. Ges. **55**, 330, 1978 (No. 6). *E*

A. J. R. de Kock: Point defect condensation in dislocation-free silicon crystals.
Semiconductor silicon 1977, Electrochem. Soc. Proc. Vol. 77-2, pp. 508-520. *E*

F. Kools: Factors governing the alignment of sintered anisotropic M-type ferrites.
Ber. Dtsch. Keram. Ges. **55**, 296-297, 1978 (No. 6). *E*

F. Kools: Effect of silica during sintering of M-type ferrites.
Ber. Dtsch. Keram. Ges. **55**, 301-304, 1978 (No. 6). *E*

P. K. Larsen, N. V. Smith*, M. Schlüter*, H. H. Farrell*, K. M. Ho & M. L. Cohen**** (* Bell Laboratories, Murray Hill, N.J.; ** University of California, & Lawrence Berkeley Laboratory, Berkeley, Calif.): Surface energy bands and atomic position of Cl chemisorbed on cleaved Si(111).
Phys. Rev. B **17**, 2612-2619, 1978 (No. 6). *E*

S. S. Lau (California Institute of Technology, Pasadena) **& W. F. van der Weg** (Philips Res. Labs, Amsterdam Division): Solid phase epitaxy.
Thin films — interdiffusion and reactions, ed. J. M. Poate, K. N. Tu & J. W. Mayer, pp. 433-480; Wiley, New York 1978.

R. E. van de Leest & A. Geven: Grain boundary effects with solid-state ion-selective electrodes.
J. electroanal. Chem. **90**, 97-104, 1978 (No. 1). *E*

F. H. de Leeuw, R. van den Doel & J. M. Robertson: The dynamical behavior of magnetic domain walls and magnetic bubbles in single-, double-, and triple-layer garnet films.
J. appl. Phys. **49**, 768-783, 1978 (No. 2). *E*

M. Lemke & W. Hoppe: Glass-cavity-stabilised Gunn oscillator.
IEE J. Microw. Opt. Acoust. **2**, 129-133, 1978 (No. 4). *H*

G. M. Loiacono & W. N. Osborne: Crystal growth from solution using cylindrical seeds.
J. Crystal Growth **43**, 401-405, 1978 (No. 4). *N*

T. A. M. Maas (Philips Philite and Metalware Manufacturing Division, Eindhoven): Optimalization of processing conditions for thermosetting polymers by determination of the degree of curing with a differential scanning calorimeter.
Polymer Engng. Sci. **18**, 29-32, 1978 (No. 1).

J. J. Maes, A. G. van Nie & G. B. A. Hut (Philips Elcoma Division, Eindhoven): Thick photoresist patterns for selective electroplating.
Microelectronics and Reliability **17**, 325-332, 1978 (No. 2). *E*

G. M. Martin & D. Bois (INSA, Villeurbanne): A new technique for the spectroscopy of deep levels in insulating materials. Application to the study of semi-insulating GaAs.
Semiconductor characterization techniques, Electrochem. Soc. Proc. Vol. 78-3, pp. 32-42, 1978. *L*

R. Memming: Reactions on semiconductor electrodes.
Topics in surface chemistry, ed. E. Kay & P. S. Bagus, pp. 1-28; Plenum Press, New York 1978. *H*

D. Meyer-Ebrecht: Images for medical diagnoses.
J. clin. Engng. **3**, 125-136, 1978 (No. 2). *H*

J. Michel, E. Fabre, M. Mautref & P. Mas: Reflectivity analysis on silicon solar cells.
13th IEEE Photovoltaic Specialists Conf., Washington D.C. 1978, pp. 1244-1248. *L*

A. R. Miedema: Surface energies of solid metals.
Z. Metallkunde **69**, 287-292, 1978 (No. 5). *E*

A. E. Morgan & H. W. Werner: Molecular versus atomic secondary ion emission from solids.
J. chem. Phys. **68**, 3900-3909, 1978 (No. 8). *E*

K. W. Moulding & G. A. Wilson: A fully integrated five-gyrator filter at video frequencies.
IEEE J. **SC-13**, 303-307, 1978 (No. 3). *R*

P. Mürau & B. Singer: The understanding and elimination of some suspension instabilities in an electrophoretic display.
J. appl. Phys. **49**, 4820-4829, 1978 (No. 9). *N*

F. W. Oliver, K. W. West, R. L. Cohen (all with Bell Laboratories, Murray Hill, N.J.) **& K. H. J. Buschow:** Mössbauer effect of ^{151}Eu in $EuNi_5$, $EuMg_2$ and their hydrides.
J. Physics F **8**, 701-707, 1978 (No. 4). *E*

K. Pasedach: On the reduction of storage space and processing time for matrices.
Z. angew. Math. Mech. **58**, T 430-432, 1978 (No. 7). *H*

J.-R. Périlhou: Une microcaméra de télévision en couleur de 2,5 mm de diamètre.
Onde électr. **58**, 319-323, 1978 (No. 4). *L*

D. Polder & K. Weiss: Remarks on the relaxation-time approximation.
Phys. Rev. A **17**, 1478-1482, 1978 (No. 4). *E*

G. Pouvesle: Acta Electronica and Georges-Albert Boutry.
Acta Electronica **21**, 109-110 *(in French)*, 111-112 *(in English)*, 1978 (No. 2). *L*

J. E. Ralph & M. J. Plummer: Charge storage and photoconductivity of PbO powder layers.
Appl. Phys. Letters **32**, 744-746, 1978 (No. 11). *R*

A. Rocher & G. Jacob: Observations préliminaires faites par microscopie électronique sur du nitrure de gallium.
J. Physique **39**, C2/121-122, 1978 (Colloque C2). *L*

S. K. Salmon: Surface acoustic wave oscillators at u.h.f.
Conf. on Radio receivers and associated systems, Southampton 1978 (IERE Conf. Proc. No. 40), pp. 101-106. *R*

S. K. Salmon: A 1.1 to 2.4 GHz transistor amplifier with low noise figure and good power match.
Conf. on Radio receivers and associated systems, Southampton 1978 (IERE Conf. Proc. No. 40), pp. 189-194. *R*

C. W. J. Schiepers (Institute for Perception Research, Eindhoven): Global attributes in visual word recognition: contour perception of three-letter strings.
Acta psychol. **42**, 205-223, 1978.

C. Schiller & J. P. Gowers: Lattice mismatch effects in the vapour phase epitaxial growth of ternary III-V semiconductors.
Acta Electronica **21**, 167-178, 1978 (No. 2). *L, R*

H. Schomberg: Parallel computers and parallel numerical methods.
Z. angew. Math. Mech. **58**, T 439-440, 1978 (No. 7). *H*

H. Schomberg: Monotonically convergent iterative methods for nonlinear discrete elliptic boundary value problems.
Z. angew. Math. Mech. **58**, T 440-441, 1978 (No. 7). *H*

H. R. Sethi & D. E. Radley (University of Sheffield): A new method for the synthesis of electron flow systems.
Int. J. Electronics **44**, 473-480, 1978 (No. 5). *R*

A. P. Severijns, F. A. Staas & W. A. Cense: An improved He^3-He^4 mixing chamber for single-cycle experiments.
Cryogenics **18**, 87-89, 1978 (No. 2). *E*

R. J. Seymour, B. J. Fitzpatrick & R. N. Bhargava: Optically pumped stimulated emission in ZnSe.
IEEE J. **QE-14**, 462-463, 1978 (No. 7). *N*

M. Sintzoff: On language design for program construction.
ALGOL Bull. No. 42, 74-84, 1978. *B*

W. T. Stacy & J. J. Daniele: Interfacial elastic strains in (Al, Ga)As/GaAs heterostructures grown by liquid phase epitaxy.
J. Crystal Growth **44**, 23-28, 1978 (No. 1). *E, N*

W. T. Stacy, A. B. Voermans & H. Logmans: Orthorhombic magnetic anisotropy in garnet epitaxial films containing europium.
J. appl. Phys. **48**, 4634-4639, 1977 (No. 11). *E*

A. L. N. Stevels: Ce^{3+} luminescence in hexagonal aluminates containing large divalent or trivalent cations.
J. Electrochem. Soc. **125**, 588-594, 1978 (No. 4). *E*

S. Stotz: Field dependence of the electrophoretic mobility of particles suspended in low-conductivity liquids.
J. Colloid Interface Sci. **65**, 118-130, 1978 (No. 1). *A*

P. J. M. van der Straten, R. Metselaar (both with Eindhoven University of Technology) **& H. D. Jonker:** Flux growth of $ZnGa_2O_4$ single crystals.
J. Crystal Growth **43**, 270-272, 1978 (No. 2). *E*

M. Tasto & H. Schomberg: Object reconstruction from projections and some non-linear extensions.
Pattern recognition and signal processing, ed. C. H. Chen, pp. 485-503; Sijthoff & Noordhoff, Alphen aan den Rijn 1978. *H*

B. Tell & C. van Opdorp: Capacitance spectroscopy of degraded GaAsP light-emitting diodes.
J. appl. Phys. **49**, 2973-2977, 1978 (No. 5). *E*

A. Thayse: Taylor expansions of Boolean functions and test functions for switching networks.
Philips J. Res. **34**, 143-146, 1979 (No. 3/4). *B*

A. Thayse, M. Davio & J.-P. Deschamps: Optimization of multivalued decision algorithms.
Proc. 8th Int. Symp. on Multiple-valued logic, Rosemont, Ill., 1978, pp. 171-178. *B*

J. B. Theeten & F. Hottier: Méthode d'analyse in situ de l'interface de croissance.
Acta Electronica **21**, 231-238, 1978 (No. 3). *L*

J. B. Theeten, F. Hottier & J. Hallais: On-time determination of the composition of III-V ternary layers during VPE growth.
Appl. Phys. Letters **32**, 576-578, 1978 (No. 9). *L*

G. E. Thomas & B. R. de Koning: Rotational and vibrational state population distributions in excited CH radicals sputtered from a chemisorption layer on silicon.
Chem. Phys. Letters **55**, 418-423, 1978 (No. 3). *E*

W. Tolksdorf & F. Welz: Crystal growth of magnetic garnets from high-temperature solutions.
Crystals — growth, properties, and applications, Vol. 1, ed. C. J. M. Rooijmans, pp. 1-52; Springer, Berlin 1978. *H*

M. J. Underhill: Oscillator noise limitations.
Proc. Conf. on Electromagnetic compatibility, Guildford 1978 (IERE Conf. Proc. No. 39), pp. 109-118. *R*

M. J. Underhill: Comparison of the noise performance of some oscillators for tunable receivers.
Conf. on Radio receivers and associated systems, Southampton 1978 (IERE Conf. Proc. No. 40), pp. 237-252. *R*

M. J. Underhill, P. A. Jordan & M. Sarhadi (Chelsea College, London): Fast digital frequency synthesiser.
Electronics Letters **14**, 342-343, 1978 (No. 11). *R*

M. J. Underhill, M. Sarhadi* & C. S. Aitchison* (* Chelsea College, London): Fast-sampling frequency meter.
Electronics Letters **14**, 366-367, 1978 (No. 12). *R*

R. Veilex: Introduction *(to issues on Vapour phase epitaxy of semiconductors).*
Acta Electronica **21**, 113-115, 1978 (No. 2). *(In English and in French.)*

J. D. B. Veldkamp & N. Hattu: On the fracture toughness of brittle materials.
Philips J. Res. **34**, 1-25, 1979 (No. 1/2). *E*

C. H. F. Velzel: Aberration balancing for grating mountings with large aberrations.
J. Opt. Soc. Amer. **68**, 38-41, 1978 (No. 1). *E*

G. Verspui: The preparation and application of pyrolytic silicon carbide.
Proc. 6th Int. Conf. on Chemical vapor deposition, Atlanta 1977 (Electrochem. Soc. Proc. Vol. 77-5), pp. 366-381. *E*

H. Verweij, H. van den Boom & R. E. Breemer: Raman spectroscopic study of the reactions in a potassium carbonate-silica glass-forming batch.
J. Amer. Ceramic Soc. **61**, 118-121, 1978 (No. 3/4). *E*

L. L. M. Vogten (Institute for Perception Research, Eindhoven): Simultaneous pure-tone masking: the dependence of masking asymmetries on intensity.
J. Acoust. Soc. Amer. **63**, 1509-1519, 1978 (No. 5).

L. L. M. Vogten (Institute for Perception Research, Eindhoven): Low-level pure-tone masking: a comparison of 'tuning curves' obtained with simultaneous and forward masking.
J. Acoust. Soc. Amer. **63**, 1520-1527, 1978 (No. 5).

Q. H. F. Vrehen & H. M. J. Hikspoors: Parametric generation in cesium vapor by nearly resonant two-photon pumping.
Optics Comm. **18**, 113-114, 1976 (No. 1). *E*

L. Vriens & M. Adriaansz: Laser diagnostics in high pressure, high temperature systems.
Optics Comm. **18**, 187, 1976 (No. 1). *E*

R. Wells: SSB for VHF mobile radio at 5 kHz channel spacing.
Conf. on Radio receivers and associated systems, Southampton 1978 (IERE Conf. Proc. No. 40), pp. 29-36. *R*

C. Werkhoven: Disclosure of dislocation loops and dislocation dipoles in GaP by chemical etching.
J. Electrochem. Soc. **125**, 671-673, 1978 (No. 4). *E*

R. Wernicke: The influence of kinetic processes on the electrical conductivity of donor-doped $BaTiO_3$ ceramics.
Phys. Stat. sol. (a) **47**, 139-144, 1978 (No. 1). *A*

R. Wernicke: Influence of microstructure on the electrical properties of intergranular capacitors.
Ber. Dtsch. Keram. Ges. **55**, 356-358, 1978 (No. 7). *A*

J. S. C. Wessels & M. T. Borchert: Polypeptide profiles of chlorophyll · protein complexes and thylakoid membranes of spinach chloroplasts.
Biochim. biophys. Acta **503**, 78-93, 1978 (No. 1). *E*

P. D. White & R. Stevens: Surface acoustic wave resonator filters.
Conf. on Radio receivers and associated systems, Southampton 1978 (IERE Conf. Proc. No. 40), pp. 93-100. *R*

G. Will, M. O. Bargouth (both with Universität Bonn) **& K. H. J. Buschow:** Magnetic properties of $GdMg_3$ and $TbMg_3$.
Rare earths and actinides 1977, Int. Conf. Durham, England (Inst. Phys. Conf. Ser. No. 37), pp. 273-277; 1978. *E*

G. Will*, K. H. J. Buschow & V. Lehmann* (* Universität Bonn): Magnetic properties and neutron diffraction of TbB_2.
Rare earths and actinides 1977, Int. Conf. Durham, England (Inst. Phys. Conf. Ser. No. 37), pp. 255-261; 1978. *E*

J. P. Woerdman: New Doppler-free two-photon lines in high-density sodium vapor.
Optics Comm. **18**, 223-224, 1976 (No. 2). *E*

P. Zandveld: Crystal damage and the properties of implanted *p-n* junctions in silicon.
Solid-State Electronics **21**, 721-727, 1978 (No. 5). *E*

H. Zijlstra: Some economic aspects of rare-earth permanent magnets.
Rare earths and actinides 1977, Int. Conf. Durham, England (Inst. Phys. Conf. Ser. No. 37), pp. 327-333; 1978. *E*

M. de Zwart: Distortion of the cholesteric planar texture in liquid crystals with a negative dielectric anisotropy.
J. Physique **39**, 423-431, 1978 (No. 4). *E*

Contents of Philips Telecommunication Review **37**, No. 1, 1979:

P. Bikker & W. J. Ham: A microprocessor-controlled communications receiver (pp. 1-10).
N. Bakker: In-line measurement of moisture in sealed IC packages (pp. 11-19).
J. van Duuren & W. J. A. Vonk: AEROPP/DS-714 data switching systems for civil aviation (pp. 20-33).
D. Muilwijk: Tamed Frequency Modulation — a bandwidth-saving digital modulation method, suited for mobile radio (pp. 35-49).

Contents of Philips Telecommunication Review **37**, No. 2, 1979:

J. Gietema: Radar display developments (pp. 53-63).
J. Gietema: A digital scan converter for radar pictures (pp. 64-70).
J. Michaelsen & S. Poulsen: Use of facilities in an EBX 8000 PABX (pp. 71-80).
M. M. Jung: Random overflow of calls between subgroups in a full-availability group of lines (pp. 81-94).

Contents of Electronic Components and Applications **1**, No. 2, 1979:

A. A. J. Franken: 2/3 inch 'Plumbicon' tube with enhanced performance (pp. 71-77).
J. van Straaten: Digital control of radio and audio equipment, Part 1 — Frequency measurement and display system for a.m. and f.m. radios (pp. 78-89).
E. D. van Veldhuizen: The 8X300 — a high-speed control-oriented microprocessor (pp. 90-101).
P. G. J. Barten & J. Kaashoek: 30AX self-aligning 110° in-line colour TV display (pp. 102-108).
C. Franx: Solar panels for terrestrial applications (pp. 109-124).
BZW10 transient suppression bridge for telephones (pp. 125-127).

Contents of Electronic Components and Applications **1**, No. 3, 1979:

S. Sliedregt: Reliability proven: BDW55/56 gold-gold medium-power transistors (pp. 135-148).
TDA1011: versatile 2 to 6 W audio amplifier (pp. 149-151).
A. Lieders: Single-phase rectifier circuits with CR filters, Part 1 — Theory (pp. 153-163).
P. G. Noble: Understanding thyristor and triac data (pp. 164-181).
M. Herrmann & J. Wölber: The P^2CCD 500B linear imager (pp. 183-191).

Volume 39, 1980, No. 3/4 pages 65-124 Published 18th November 1980

Supplement to Philips Technical Review
Beilage der Philips Technischen Rundschau
Bijlage van Philips technisch Tijdschrift

PHILIPS

October 1980

Recent United States Patents

Abstracts from patents that describe inventions from the following research laboratories that form part of or cooperate with the Philips group of companies:

Philips Research Laboratories, Eindhoven, The Netherlands	*E*
Philips Research Laboratories, Redhill, Surrey, England	*R*
Laboratoires d'Electronique et de Physique Appliquée, 3 avenue Descartes, 94450 Limeil-Brévannes, France	*L*
Philips GmbH Forschungslaboratorium Aachen, Weißhausstraße, 51 Aachen, Germany	*A*
Philips GmbH Forschungslaboratorium Hamburg, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany	*H*
Philips Research Laboratory Brussels, 2 avenue Van Becelaere, 1170 Brussels (Boitsfort), Belgium	*B*
Philips Laboratories, N.A.P.C., 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, USA	*N*

4 187 139

Growth of single crystal bismuth silicon oxide

J. C. Brice *R*
O. F. Hill
R. G. Pratt

Method of growing single crystal of bismuth silicon oxide from a melt of $Bi_xSiO_{1.5\,x+2}$ in an oxygen containing atmosphere. At the start of crystallization, x has a value of 11.55 to 11.82 or 12.35 to 14.00. The value of x does not have a spread of more than 0.025 throughout the crystal.

---

4 188 433

Record carrier in disk form having cover layers

R. Dijkstra *E*
A. J. M. van den Broek
G. J. M. Lippits

The invention relates to a disk-shaped record carrier, for example a video plate, which is provided on one side or on both sides with a radiation-reflecting surface structure in the form of an information track consisting of blocks and pits on which a reflecting layer of metal is provided. The plate can be read optically. According to the invention the reflecting metal layer is provided with a cover layer which is formed by a transparent plate which is connected to the metal surface by means of a radiation-curable lacquer. The lacquer is preferably a U.V. curing lacquer and contains a protic mixture of acrylic acid esters. The transparent plate is also preferably manufactured from a transparent synthetic resin and has a thickness of 200 μm.

---

4 188 562

Color display tube and method of manufacturing such a color display tube

K. J. van Oostrum *E*

A color display tube having a quadrupole post-focusing color selection electrode comprising a metal plate having a very large number of apertures, one side of the plate being covered with an insulating film on which a metal film is provided. The insulating film and metal film extend on two opposite portions of the wall of each aperture so that the plate and metal film form a pair of electrodes for producing an electron lens in the apertures.

4 188 638

Color television signal having color-difference signals alternating between two carriers

M. R. de Haan *E*

A transmission system for a color television signal, in particular a system for recording on and reproducing from a record carrier. The chrominance signal is divided into two separate color signals, which in a line-sequentially alternating fashion are modulated on two separate color carrier waves. The transmitted or recorded color television signal can simply be transformed into both a standard PAL and a standard SECAM signal. Owing to the line sequential alternation of the color signal, color errors which may arise owing to a mutual difference of the transfer characteristics for the two color signals are automatically compensated for in the case of reproduction via a PAL receiver.

---

4 188 640

Apparatus for making laminar radiograms

J. Dittrich *H*
J. Heinzerling
D. E. Meijer
R. Möllendorf
F. Wolf

In tomosynthesis, as is known, a series of individual images taken in different orientations, are superimposed. By shifting the individual images relative to each other the position of the layer which is sharply reproduced can be changed. However, there may still be unsharpness in said layer, in the case of image distortion during recording of the individual images, for example owing to the curvature of the input screen. This image distortion can hardly be compensated for, because it greatly depends on the projection angle. The invention provides a simple method of compensating for such image distortion. The image intensifier is then arranged and steered so that its optical axis is always parallel to the central ray. The distortion caused by the image intensifier curvature is then independent of the projection angle and can be eliminated in a comparatively simple manner. Steering the image intensifier in this way gives rise to additional distortion, because the planes, of which subsequently laminar radiograms are made by superposition, and the optical axis of the image intensifier are no longer perpendicular to each other. However, this is a simple geometric distortion which can be eliminated in a simple manner.

4 188 791

Piston-centering system for a hot gas machine

J. Mulder E

A free piston hot gas reciprocating machine having one piston surface which varies the volume of a working space, while its other surface bounds a buffer space of constant pressure, and a control mechanism formed by an auxiliary cylinder and an auxiliary piston which is movable therein. The control mechanism maintains a given central position of the free piston by instantaneously opening a connection between the buffer space and the auxiliary cylinder space.

---

4 189 342

Semiconductor device comprising projecting contact layers

H. G. Kock E

A microwave semiconductor device, for example, Gunn effect devices, avalanche diodes and varactors, having a mesa-like semiconductor body which is provided on both sides with respective grown metal contact layers. The upper side of the semiconductor body is provided with a beam lead which projects from the upper side. The metal contact layer of lower side of the semiconductor body is a grown heat sink or, as in the varactors, a beam lead.

---

4 189 393

Heat storage material comprising lithium chlorate-trihydrate and a nucleating agent

J. Schröder A
K. Gawron

A heat storage material comprising lithium chlorate-trihydrate as the heat storage medium and potassium perchlorate and/or sodium, potassium or barium hexafluorosilicate as the nucleating agent to reduce supercooling.

---

4 189 394

Heat storage material comprising calcium chloride-hexahydrate and a nucleating agent

J. Schröder A
K. Gawron

The utility of calcium chloride-hexahydrate as a heat storage material is improved when barium carbonate, strontium carbonate, barium fluoride, barium fluoride-hydrofluoride and/or strontium fluoride is used as a nucleating agent to prevent supercooling.

---

4 190 861

Method and arrangement for redundancy-reducing picture coding

P. Lux H

With known methods for predictive decorrelation this decorrelation only applies to one or more previously scanned elements of the same picture line. In accordance with the invention a picture to be scanned is divided into sub-pictures wherein a number of picture lines equal to the number of lines comprised in a sub-picture is intermediately stored in shift registers, a processing circuit being connected to the last stages of these shift registers, which processing circuit multiplies the values of the picture elements of one sub-picture with the same number of coefficients stored in the processing device and which adds the products. In a preferred embodiment the difference values obtained in this manner are also decorrelated in a second stage, the matrix of the difference values being divided into difference value sub-pictures which are shifted, preferably diagonally, relative to the original sub-pictures. A considerably larger surrounding area of each picture element is thus involved in the prediction.

4 190 871

Magnetic converter having a magnetoresistive element

A. Walraven E

A magnetic converter having a magnetoresistive element, the resistance of which is measured by means of an alternating current having a frequency which is larger than $8\varrho/\mu D$, where ϱ is the resistivity, μ the magnetic permeability and D the cross-section of the element so that the current distribution is concentrated at the edges of the element due to the resultant lateral skin effect.

---

4 191 956

Device for the display of data on a display apparatus

H. H. H. Groothuis E

A device for displaying page data on a television monitor, having a foreground memory and a background memory. The foreground memory has a capacity of one page of text. The background memory has serial sub-memories connected in parallel, each sub-memory containing part of a character, so that together the sub-memories contain the relevant character within a series of characters. The characters become serially available on the outputs of the sub-memories for transfer, to the foreground memory. The sub-memories may be magnetic bubbles. Each of said sub-memories preferably is a single loop so that intermediate control during the transfer of a page of data is very simple. The foreground memory is preferably a random access memory which can be addressed. The output of the memory has connected to it a counter with sub-counting cycles having a length of one line on the display apparatus in order to form the lines to be displayed thereon.

---

4 193 033

Quadrature transposition stage

J. O. Voorman E

Quadrature transposition stage has two signal channels, each containing a filter. A gyrator circuit is coupled between the filters. The result is that the two filters have exactly the same transfer function which reduces signal distortion in such circuits as an FSK demodulator.

---

4 193 091

Optical videodisc read unit with tracking and focusing wobblers

W. J. Kleuters E
G. E. van Rosmalen
M. P. M. Bierhoff
K. A. Immink

An optical read unit for reading radiation-reflecting video discs comprises a semiconductor laser which with the aid of a scan wobbler and a focusing wobbler is moved at high frequency in two directions relative to the video disc, a radial direction for tracking purposes and a direction normal to the disc for the purpose of automatic focusing. The semiconductor diode laser is mounted on a metal radiation source carrier which is connected to a heat sink for carrying off the heat with a low temperature gradient, and is driven by piezo-electric plates.

---

4 194 133

Charge coupled circuit arrangements and devices having controlled punch-through charge introduction

J. M. Shannon R

In a CCD introduction of charge with good linearity and low signal input noise sensitivity is obtained by the control of punch-through of a depletion region associated with an input storage site to a rectifying barrier bounding a region in the semiconductor body separated from the CCD channel and forming a source of charge carriers to be stored and transported. In one form the source of charge carriers is an opposite conductivity type substrate on which is present a region of the one conductivity type in which a surface channel CCD is present. In other forms, including both surface channel CCDs and buried channel CCDs, buried layers are employed as source regions.

4 195 223

Apparatus for optically scanning a surface

U. Bergmann *H*
K. Klose

Apparatus for optically scanning a surface, the radiation path between the light source and light detector including a wedge for correcting undesired intensity variations which occur when a homogeneous non-textured surface is scanned.

---

4 195 271

Broad-band 180° phase shifter

B. Schiek *H*
J. Köhler

Disclosed is a 180° phase shifter having meander-shaped coupled lines in a predominantly planar form, such as, for example, microstrip, microslot or stripline, formed by a $\lambda/2$ long meander disposed opposite to a $\lambda/4$ long meander.

---

4 197 121

Method of making electrophotographic images with a uniform exposure step

W. Eckenbach *A*

The duration of use of spectrally sensitized pigment bonding agent layers which are to be used a number of times is substantially increased by performing flood-lighting, after the development and/or transfer and prior to the recharging, with a light which is suitable per se for the image formation, but which is substantially not absorbed by the dye.

---

4 197 482

Color selection means for color display tube and method of making same

J. Koorneef *E*
A. J. der Kinderen

A color display tube having a quadrupole post-focusing electrode consisting of two crossing sets of parallel conductors, the first set of conductors consisting of wires and the second set of conductors consisting of metal stripes secured pairwise against each other and clamping the wires at the crossings, said wires being surrounded coaxially at the area of the crossings by a layer of insulation material, has the advantage that already at a comparatively low focusing voltage between the sets of conductors a sufficiently strong focusing influence is exerted on the electron beams.

---

4 198 954

Solar collector, comprising solar tracking means

R. J. Meijer *E*

A solar collector comprising a focusing reflector rotatable about an axis; an adjusting mechanism for effecting rotation of the reflector about such axis; and a solar-tracking arrangement for actuating the adjusting mechanism, the solar-tracking arrangement including a pair of sun-sensor reservoirs each containing a high-pressure working medium that thermally expands when subjected to solar radiation.

Re. 30 251

Semiconductor device comprising an insulated gate field effect transistor and method of manufacturing the same

E. Kooi *E*

Method of making an insulated gate field effect transistor is described in which the surface of a silicon semiconductor is covered in whole or in part with a layer of a masking material which masks against oxidation, such as silicon nitride. Areas of the silicon surface are exposed for the source and drain regions, leaving the oxidation mask over the future channel. When the source and drain regions have been made, as for example by diffusion, the device is subjected to oxidation, causing the growth of a thick oxide which sinks into the silicon surface where it is not masked by the oxidation mask. Among the advantages obtained are fewer precise masking steps, a flatter device surface, and reduced gate overlap of the source and drain.

---

4 199 378

Method of manufacturing a semiconductor device and semiconductor device manufactured while using such a method

J. A. A. van Gils *E*

A method of manufacturing LOCOS transistors in which base doping, emitter doping and emitter metallization are provided via the same aperture. Problems at the edge of the sunken oxide are eliminated by a two-stage doping technique so that the channel stopper diffusion in the epitaxial layer may be omitted, which presents particular advantages in the manufacture of I^2L devices.

---

4 199 621

Method of producing iron oxide photo masks

T. J. A. Popma
M. G. J. Kamminga-Schrijnemakers *E*

Method of producing photomasks of X-ray amorphous iron oxide (Fe_2O_3) by spraying a transparent substrate at a temperature between 180 and 250 °C with a solution of a hydrolysable iron salt in a polar organic solution or, in accordance with a variant of this method, at a temperature between 180 and 750 °C with a solution of a hydrolysable iron salt which also contains a salt of a metal having a greater ion diameter. If temperatures above approximately 230 °C are used, the solvent may in the latter case also consist of water only.

---

4 199 796

Tape guide arrangement for a transversely scanning magnetic recorder

B. P. Videc *E*

A transversely scanning magnetic tape recorder having a concave tape guide member which shapes and aligns the tape at a location by the rotating magnetic heads and between two cylindrical tape guides. The magnetic tape is deformed in cross-section from straight to curved and then back from curved to straight. To prevent kinks and folds in the magnetic tape as a result of this deformation, tape curving means exert pressure on the magnetic tape to increase the local degree of curvature of the magnetic tape. As a result, differences in local membrane stress in the tape cross-section at the location of the rotary magnetic head paths are reduced.

---

4 200 887

Television camera

P. M. van den Avoort *E*

A television camera in which during picture pick-up a same, non-interlaced line raster of lines of half the standard line period are processed in a standard field period. Thereafter signal processing such as a vertical aperture correction and/or an interlace-flicker reduction, takes place in an advantageous manner without a field period delay for improving the picture quality by combining the information of spatially directly adjacent lines. Thereafter signal expansion of the information from half the standard line periods to the standard line periods take place so that ultimately a video signal in accordance with the standard is obtained.

4 201 954

Gas discharge laser for generating linearly polarized radiation

J. van der Wal E
G. Bouwhuis

A gas discharge laser includes one or more birefringent layers in a multi-layer reflector that provides such a large difference in reflection of light polarized linearly in two mutually perpendicular directions that the laser generates 100% linearly polarized light in one direction only. In the other direction the reflection is insufficient to obtain stimulated emission.

---

4 204 124

Device for measuring radiation absorption in a three-dimensional object

G. Kowalski H

A device for reconstructing the absorption distribution in a three-dimensional body. The body to be examined is imaged from two groups of radiation source positions. Each group of radiation source positions is situated on an arc of a circle around the axis of rotation, and the arcs of a circle determine the dimensions of the region in which the absorption distribution is reconstructed. The detector device is disposed between the two arcs of a circle.

---

4 204 163

Minimum group pulse code modem having shape and amplitude codes

F. de Jager E

A transmission system which on the transmission side comprises a pulse-code modulation device, which is adapted for the pulse code conversion of consecutive minimum groups of two signal elements of the information signal, the pulse code modulation device comprises a shape encoder, which in each group interval, encodes by a pulse code a normalized-amplitude shape of the signal variation of the two signal elements, and furthermore comprises an amplitude encoder, which in each group interval, encodes an associated amplitude value by a pulse code. The receiver comprises a multiplier stage for multiplication of the normalized-amplitude shape, which is transmitted in pulse code, in each group interval, by the associated amplitude value.

---

4 204 177

Non-recursive digital filter with reduced output sampling frequency

L. D. J. Eggermont E

Non-recursive digital filters having an output sampling frequency which is a factor of r lower than the input sampling frequency. A number of filters are described in which for each value of r and N (N being the number of filter coefficients) the time which is required to calculate an output sample is based on the output sampling period T_u. This is achieved by using a storage device in the input circuit.

---

4 204 253

Device for generating and correcting a user program

J. G. van den Hanenberg E
F. J. de Munnik

The invention relates to a programming panel which preferably comprises a micro-processor and whereby an operator can generate or modify the contents of steps of a user program by means of a number of selection means and value input means, inter alia on the basis of indications displayed on display means for this purpose. The user program is at the disposal of a machine control system, it preferably being possible to couple the programming panel to the system by way of a bus connection. The programming panel comprises standard selection means, notably for so-termed movement primitives.

4 204 743

Connector for coupling optical fibres to an emitter or receiver of luminous energy

J.-F. Etaix L

A connector for connecting optical fibres to an emitter or a receiver comprises two connection elements which are adapted to be coupled and locked relative to each other. One of the connection elements encloses the end of the fibre bundle. The other element comprises an assembly which connects an optical imaging forming device to the emitter or receiver. Each of the elements comprises internal parts which, during the coupling, enable the end of the fibre bundle and the surface of the emitter or the receiver to be positioned relative to each other with high accuracy, in the transverse as well as in the axial direction, the emitter or receiver and the bundle thus being optically coupled by the optical device.

---

4 204 769

Cuvette for flameless atom absorption spectroscopy

B. Lersmacher A
H. Lydtin
K. Schelhas
W. F. Knippenberg

Cuvettes of carbonized laminated fabric with a surface layer of pyrolytic graphite are easy to manufacture, have a high mechanical rigidity and no memory effect and enable the operation of analysis apparatuses at normal mains voltage (220 V).

---

4 205 215

Method and device for welding in a thermally ionized gas

G. A. M. Willems E

A plasma-MIG welding system in which the power for establishing the MIG-arc is supplied to the consumable electrode at a point downstream of the non-consumable electrode.

---

4 206 374

Synchronous motor

B. H. A. Goddijn E

A synchronous motor, in particular a stepping motor, with two coaxial stator sections which are each provided with a ring coil and with two circular systems of teeth, which are shifted by half a tooth pitch relative to each other. The two stator sections are shifted by a quarter tooth pitch relative to each other. In addition, the motor comprises an axially magnetized permanent-magnetic ring which is disposed between the two stator systems, for magnetizing the two systems of teeth of the stator sections with mutually the same polarity. The two systems of teeth of each stator section are constituted by annular parts which are provided with teeth at the inner circumference. The annular parts form part of the magnetic circuit which surrounds the annular coil and are disposed concentrically in such a way that the axial distance between the teeth of both systems is at least a few times greater than the air gap between stator and rotor.

---

4 207 477

Bulk channel CCD with switchable draining of minority charge carriers

L. J. M. Esser E

The invention relates to a charge-coupled device in which the charge transport in the form of majority charge carriers takes place mainly via the bulk of a semiconductor layer of one conductivity type. The semiconductor layer has zones of the second conductivity type which do not have an electric contact but which are electrically biased by means of the isolation zone surrounding the semiconductor layer which can be connected to the zones by induction by means of the electrodes and forms a drain for charge carriers from the zones. In an embodiment the device is formed by a two-phase charge-coupled device in which the zones serve to obtain asymmetry in the system. In another embodiment the device is a series-parallel-series multiplex CCD in which the zones form isolation zones between the parallel lines.

PHILIPS TECHNICAL REVIEW

VOLUME 39, 1980, No. 5

Chromatofocusing, a new protein-separation method

L. A. Æ. Sluyterman

The chemist's understanding of the structure of protein molecules and his control of protein chemistry have made great strides in the last twenty years or so. In the sixties, for example, proteins consisting of as many as 51 or even 125 amino acids were successfully synthesized — insulin and the enzyme ribonuclease. No less remarkable are the achievements of research on recombinant DNA: the use of bacteria to produce insulin and interferon (a natural protein that enhances human resistance to disease) in quantity.

The brilliance of these achievements should not however blind us to the progress made in the basic techniques of 'mixing and separation', which are after all fundamental to all chemistry. The article below discusses a new method of protein separation.

Electrophoresis is a well-known method of separating particles that carry an electric charge — such as colloidal particles and proteins. A particularly elegant version of this method, which has been in use for the last 15 years or so, especially for the separation of proteins, is the isoelectric focusing method [1]. In this method the protein — whose molecule can carry both positive and negative charges at different sites (*fig. 1*) — is introduced into a *p*H gradient, which is created by means of an electric current. The method separates the proteins (and similar substances) by virtue of their difference in isoelectric point *p*I — the *p*H at which the complete protein molecule is electrically neutral and does not move in an electric field. The proteins are 'focused' into narrow zones, or bands, at the position in the column where the *p*H is the same as their isoelectric point. (The *p*I of each protein is then easily determined by measuring the *p*H at the centre of each band.)

We have succeeded in producing the same focusing effects by means of chromatography, using a column filled with an appropriate ion exchanger. We have called the method *chromatofocusing*. An important advantage of the method is that cooling presents no problems, even in separations of large amounts of material; this is not so for the electrophoretic method.

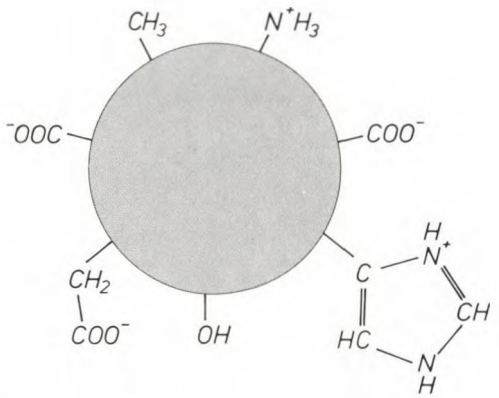


Fig. 1. Schematic representation of a protein molecule. The molecule consists of a single long chain formed by hundreds of amino acids, generally folded in a fixed configuration. Here and there the chain has a side-chain, which can carry a charge. In this way phenolic OH groups and COOH groups can release a proton, while amino and imidazol groups can take up a proton.

In chromatofocusing, a *p*H gradient must also be created. The procedure is broadly as follows. The column is filled with either an anion or a cation exchanger — in this article we shall take, by way of example, an anion exchanger. Next, a buffer with a relatively *high p*H value is passed through the column. Once this buffer and the ion exchanger — which also

Prof. Dr L. A. Æ. Sluyterman, formerly with Philips Research Laboratories, Eindhoven, is now a Visiting Professor at Eindhoven University of Technology.

[1] See for example H. Rilbe, Science Tools **23**, No. 2, p. 18, 1976.

has a buffering action — have reached their equilibrium state, the protein, dissolved in a buffer with a relatively *low* *p*H, is introduced into the top of the column. Elution then takes place with the same buffer of relatively low *p*H, thus producing the required *p*H gradient.

In this article we shall deal with the mechanism of the focusing action, and a few experiments will be described that illustrate this action and demonstrate the usefulness of the method [2].

Separation with a *p*H gradient

Simple model

The behaviour of a protein during its passage through a column in which a *p*H gradient is formed in the manner described above can be understood by reference to *fig. 2*. The figure gives the *p*H values as a function of position in the column, calculated with the aid of a simple model, after each of 19 equal parts ('aliquots') of eluent has passed. A value of 10 was chosen for the initial *p*H of the buffer in the column, and the value of 8 for the initial *p*H of the elution buffer. It is assumed that the column is divided into 10 equal sections. In addition, the volume of the ion exchanger was neglected and each aliquot of eluent was assumed to be equal to the void volume of a column section. The graph shows that the successive aliquots of eluent each travel through a slightly different range of *p*H in the column.

Let us now suppose that a protein will only shift its position in the column — and then at the same rate as the eluent — if the total charge of the protein molecule and of the ion exchanger have the same sign, and also that the protein will not be displaced at all, that is to say it will be absorbed, if the charges are of opposite sign (the 'all-or-none' model). Let us also assume that the protein molecule carried along in this way has an isoelectric point equal to 9. Upon entering the column the protein will initially travel at the same rate as the eluent. The situation changes when a position in the column is reached where the current *p*H of the buffer in the column is equal to the *p*I of the protein. For the first aliquot that position is point *a*; the protein is now retained by the (positively charged) ion exchanger. When aliquot 2 arrives at the same site, the *p*H there has meanwhile dropped to the value α. The protein of the second aliquot therefore passes this point — and the protein of the first aliquot is also set in motion — and comes to a standstill at *b*, where in this case the *p*H has also become (slightly) higher than its *p*I value. This process is repeated up to the tenth aliquot, which only just reaches a point with *p*H = 9 before it emerges from the column.

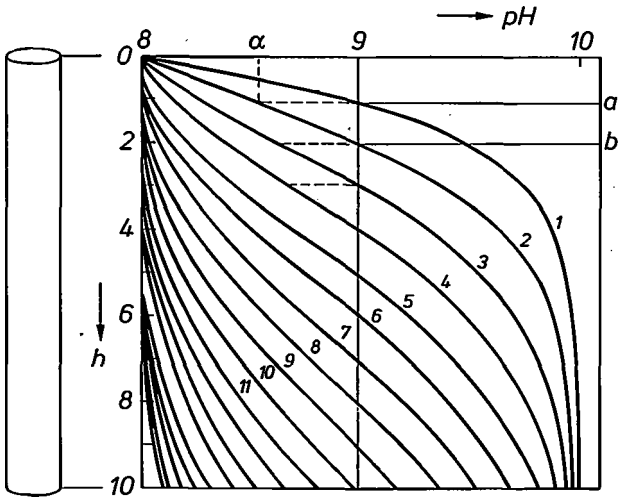


Fig. 2. The principle of protein separation by means of chromatofocusing, explained with the aid of a greatly simplified model. The graph gives calculated *p*H values of 19 equal parts ('aliquots') of a buffer solution (1 to 19; *p*H = 8) passed through a column filled with an ion exchanger that was initially equilibriated with a buffer with a different *p*H (*p*H = 10), as a function of their position *h* in the column. Addition of the aliquots produces a *p*H gradient in the column, and this gradient is slightly different after the passage of each successive aliquot. If the aliquots contain molecules of a protein, they travel along the gradient up to the point where the *p*H is equal to their isoelectric point *p*I. When they reach this point they change sign and are absorbed by the ion exchanger. For a protein molecule with *p*I = 9 in the first aliquot, this position is point *a*. When the second aliquot arrives at *a*, the *p*H there has fallen to α and the protein molecules from both aliquots move on to point *b*, and so on. In the example given here the protein molecules of the first ten aliquots bunch together and emerge from the column simultaneously.

If we now abandon the schematic representation of the eluent as consisting of a finite number of discrete aliquots, we see at once that the protein does not move in a number of separate steps but continuously. Its rate of advance will initially be high, until it reaches its isoelectric point, and then much lower. The protein that enters in the first 'aliquot' will travel for the longest time at a slow rate; the period of slow movement will be shorter for the proteins whose turn comes later. This is the qualitative explanation of the focusing effect.

It will also be clear that each different protein is focused at the site in the column whose *p*H corresponds to that of its isoelectric point.

The distribution volume

Fig. 2 also shows that a protein should be dissolved in a limited amount of liquid if the protein is to be eluted from the column in a single band. If, for example, we were to use eleven aliquots of eluent for the solution of the protein with *p*I = 9, the eleventh

[2] The theoretical models used as a basis for the optimization of the method at Philips Research Laboratories are described by L. A. Æ. Sluyterman, O. Elgersma and J. Wijdenes in J. Chromatography 150, 17 and 31, 1978.

aliquot would arrive too late. The maximum volume in which a protein may initially be dissolved to satisfy the requirement that the protein shall emerge from the column in a single band is called the *distribution volume*.

As can be shown from the graph, the magnitude of the distribution volume depends on the *p*I of the protein. In the extreme case where the *p*I of the protein is equal to the initial *p*H of the buffer liquid in the column, the protein is not retarded by the exchanger, so there is no focusing effect. The elution volume is then equal to the void volume of the column and the distribution volume is zero. As the *p*I of the protein decreases, the retarding action of the column increases, and the elution volume can be larger in relation to the void volume of the column (while preserving the condition that the protein should emerge from the column as a single band); this means that the distribution volume increases. As a first approximation, the distribution volume is equal to the elution volume less the void volume of the column.

Detailed model

The model used so far has assumed that the protein is focused in infinitely narrow bands. Let us now consider the focusing process with the aid of a more detailed model, bearing in mind that the ion exchanger consists of grains that occupy a certain volume and are capable of absorbing or desorbing ions. For an anion exchanger these ions will mainly be negative and most of the positive charge will be fixed and non-exchangeable; for a cation exchanger the opposite applies. These fixed charges give rise to a *Donnan equilibrium*: the fixed nature of part of the charge and the need to maintain electroneutrality give rise to an ion distribution such that the *p*H value in the stationary phase, pH_s, and in the mobile phase, pH_m, is not the same. (In a column filled with anion exchanger, pH_m is lower than pH_s.) The acidity measured in the emerging band containing the protein is therefore not exactly equal to *p*I [2].

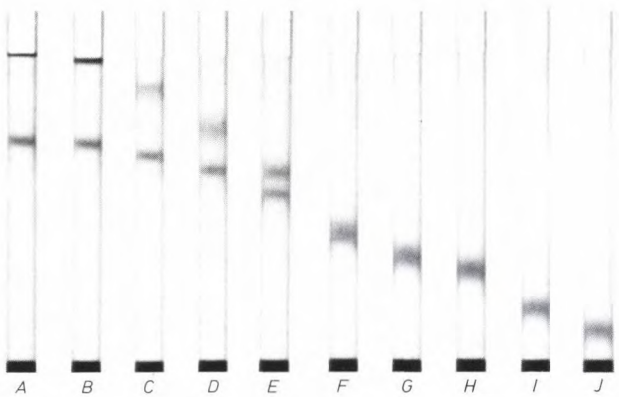


Fig. 3. Two samples of myoglobin introduced at different times into a column, filled with the anion exchanger DEAE-Bio-Gel A and eluted with ampholine, merge to form a single band within 113 minutes. Photographs *A-J* were made 0, 2, 9, 18, 32, 55, 69, 75, 98 and 113 minutes after introducing the second myoglobin sample.

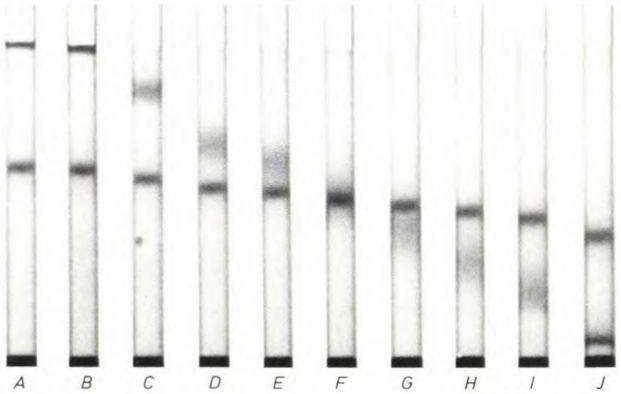


Fig. 4. An experiment similar to that in fig. 3, but this time with cytochrome c (*p*I about 10) and myoglobin (*p*I about 9). The myoglobin was applied first and after 19 minutes elution it is overtaken and passed by the cytochrome c. Photographs *A-J* were made 0, $1\frac{1}{2}$, 7, 13, 15, 19, 23, 27, 31 and 43 minutes after introducing the cytochrome c. The band of the cytochrome sample first becomes broader, owing to the 'labyrinth effect', and in a later phase it narrows, because of the focusing action. (This effect is not apparent in fig. 3 because the sample last introduced joined the first sample during the narrowing process.)

Furthermore, the fact that we are now no longer neglecting the volume of the ion exchanger leads to a bandwidth that differs from zero; when two protein molecules with the same *p*I, e.g. 9, are introduced simultaneously into the column, one molecule will travel a shorter distance between the grains of the ion exchanger than another. The competition between this 'labyrinth effect' and the focusing action during the separation results in an initial broadening of the band, later followed by a narrowing.

So far we have confined ourselves to separations with a *p*H gradient produced by the *internal* method. It is also possible to obtain the required *p*H gradient by the '*external*' method. First, the proteins are introduced into the top of the column, and they are dissolved in a buffer that has a relatively *high* *p*H (not a *low* *p*H as in the internal method). Elution next takes place with a buffer mixture whose ratio is gradually changed, so that the *p*H of the liquid introduced into the column gradually decreases.

An important requirement in a method of protein separation is that the proteins should not be unduly exposed to extreme conditions. It is particularly important, for example, that the proteins should not be exposed for too long to *p*H values that differ considerably from their *p*I value.

The natural environment for proteins usually has a *p*H value of about 7. It should not be concluded from this that all proteins presumably have a *p*I value of about 7; many proteins do in fact have *p*I values very different from 7. Such proteins, however, can always stand up well to extreme *p*H values. In the method using an internally produced *p*H gradient the proteins are dissolved in a buffer with a relatively low *p*H, in the neighbourhood of 7. Proteins with a relatively low *p*I are therefore not subjected to extreme conditions in this method. The only proteins that will be subjected

to extreme *p*H values for some time, i.e. during solution in the elution buffer and introduction into the column, are those that have a high *p*I. As already noted, however, such proteins can stand up to such *p*H values.

In the external method the proteins are all initially absorbed, at a relatively high *p*H, by the ion exchanger, and each protein is not desorbed and transported until the part of the eluent has arrived whose *p*H corresponds to its *p*I value. This means that in this case proteins with a relatively low *p*I will be exposed for a long time to high *p*H values. In the limiting case of the protein that 'comes last', a column length of eluent with *p*H values different from its *p*I has passed the protein before it is desorbed. In general, therefore, the internal method is preferable to the external one.

Experiments

Two experiments that demonstrate the focusing action will now be described.

In the first experiment we applied two samples of the same reddish-brown myoglobin to a column at different times, and then photographed the column immediately after applying the second sample and again 2, 9, 18, 32, 55, 69, 75, 98 and 113 minutes later (see pictures *A* to *J* in *fig. 3*).

In the second experiment we used the red-coloured protein cytochrome c as well as myoglobin. Cytochrome c has a larger *p*I than myoglobin and will therefore move faster through a column containing an anion exchanger than myoglobin. The cytochrome c was added some time after the myoglobin, and the series of photographs in *fig. 4* show very clearly how the cytochrome overtakes the myoglobin and passes it. It can also clearly be seen that the cytochrome band first becomes broader and then narrower.

We also performed experiments to compare chromatofocusing with the electrophoretic method. The elution patterns in *figs 5* and *6* show that our chromatographic method produces results that are at least as good as those obtained with electrophoresis.

In the chromatofocusing separation of two (or more) proteins it is necessary to take account of the distribution volume of both proteins. If the protein that emerges from the column first is applied to the column in too great a quantity of elution buffer — i.e. in a quantity greater than the distribution volume of that protein — then the focusing will not take place in the correct way.

Finally, an experiment will be described that we carried out to determine the distribution volume of a protein and to compare it with the theoretical value. We based this value on the detailed model, from which it can easily be shown that the distribution volume is equal to the elution volume less the total 'accessible' volume of the column (or the total volume of the column less the volume occupied by the molecules of the ion exchanger).

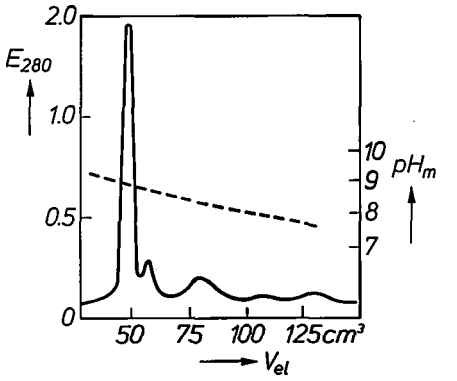


Fig. 5. The result of a purification of sperm-whale myoglobin, performed by means of chromatofocusing. The graph shows E_{280}, the extinction measured at 280 nm, as a function of the elution volume V_{el}. After the addition of 50 cm³ of eluent the main component emerges from the column. The dashed line gives the *p*H value, pH_m, of the emergent fluid.

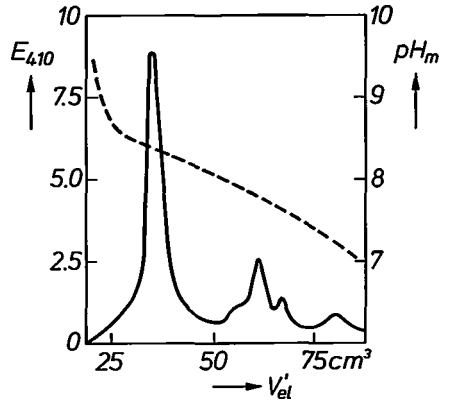


Fig. 6. Result of a purification of the same myoglobin as in fig. 5, now performed by means of isoelectric focusing. The extinction was determined this time at 410 nm.

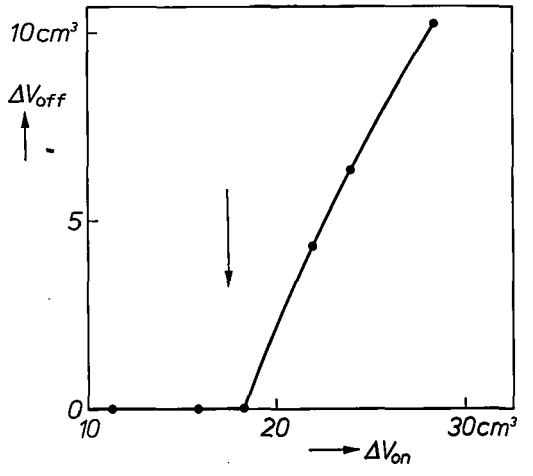


Fig. 7. The difference ΔV_{off} in buffer volume required for two separately introduced samples of myoglobin to emerge from the column, as a function of the volume difference ΔV_{on}, the elution volume passed through the column between the two applications of the proteins. The arrow indicates the calculated value of the distribution volume. It agrees well with the experimental value, i.e. the value of ΔV_{on} at which ΔV_{off} is greater than zero.

As shown in *fig. 7*, we have plotted the difference ΔV_{off} in elution volume of two protein samples, introduced at different times, as a function of the initial distance between their positions, expressed as the amount of elution buffer ΔV_{on} passed through the column between the two applications of the samples. As long as ΔV_{on} is smaller than the distribution of the protein last applied, the difference ΔV_{off} in elution volume is zero and both protein bands emerge simultaneously from the column. As soon as ΔV_{on} is greater than this distribution volume, ΔV_{off} is no longer zero and the two bands do not emerge from the column simultaneously. The value ΔV_{on} at which ΔV_{off} is greater than zero is the distribution volume. The arrow indicates the distribution volume calculated in this way. As can be seen, there is good agreement between the measured and calculated values.

Summary. When protein molecules (and other molecules or particles carrying a variable charge) pass through an appropriate *p*H range, their charge changes sign when the *p*H value in the range is reached that corresponds to their isoelectric point (*p*I). In the protein-separation method described here, called chromatofocusing, the protein molecules travel through such a *p*H range in a chromatographic column filled with an ion exchanger. On charge reversal the protein molecules are temporarily absorbed by the ion exchanger. This reduces their rate of travel as soon as they enter the part of the gradient where the *p*H is equal to their *p*I value, and this has the effect of focusing the protein molecules at a corresponding position in the column. The separation results achieved with this method are at least as good as those obtained with the widely used isoelectric focusing method based on electrophoresis. In the chromatographic method, unlike isoelectric focusing, there are no cooling problems in large-scale separations.

A method of measuring mechanical stresses in passivation layers

A. G. van Nie

Integrated circuits and other semiconductor devices for some applications are coated with a passivation layer before being encapsulated. The primary object of this layer is to protect the device from atmospheric effects — hence the name — but it also serves to provide protection from mechanical damage during encapsulation. An expensive hermetically sealed encapsulation is not necessary for semiconductor devices that have a good passivation layer, even for applications reguiring a very long service life. The application of the passivation layer is the last stage in the production process.

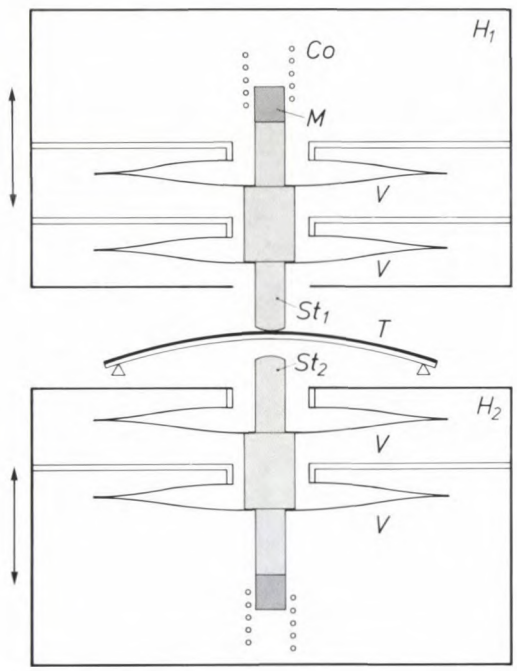


Fig. 1. Principles of the measurement of mechanical stress in the passivation layer of an IC with the aid of two spring-pressure gauges. The stress is determined by measuring the force required to straighten a glass strip, coated with a passivation layer, which is bowed by the mechanical stress in this layer. *T* test sample (supported at three points). $H_{1,2}$ housings of spring-pressure gauges; these can be moved independently of each other along a vertical rail. *V* leaf springs, two for each gauge. $St_{1,2}$ pins. The displacement of the pin by an external force can be measured electrically by means of the rods of magnetic material *M* and the coils *Co*. In the initial position the end of St_2 is located just above the plane through the support points. During the measurement, H_1 is moved downwards until the lower gauge indicates that the end of St_2 has reached this plane.

Ir A. G. van Nie is with Philips Research Laboratories, Eindhoven.

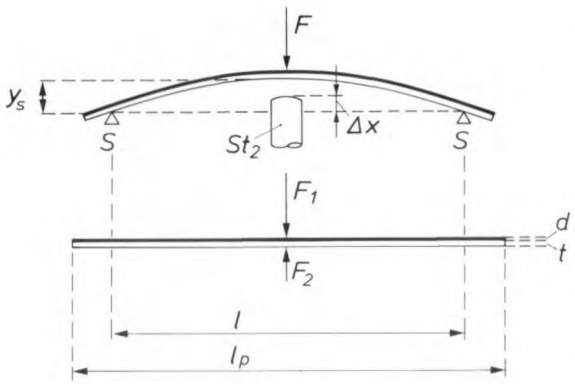


Fig. 2. Principle of calculating the relation between the compressive stress in a passivation layer and the forces F_1 and F_2 measured by the spring-pressure gauge when the sample has been straightened. The curved strip is assumed to be an arc of a circle. *S* support points. *l* distance between the support points, chord of the arc. y_s deflection (sagitta of the arc). Δx position of the end of St_2 (fig. 1) in the initial position, with respect to the plane through the support points. l_p and *t* length and thickness of the sample. *d* thickness of the passivation layer.

A material now commonly used for passivation layers is silicon nitride, which is applied by means of a plasma containing SiH_4 and NH_3. In this process, known as Plasma Enhanced Chemical Vapour Deposition (PECVD), good layers can be deposited at a fairly low temperature (about 300 °C). To prevent the occurrence of pinholes and microcracks in the layer, the process parameters are adjusted in such a way that a compressive stress is set up in the layer. This requires care, because if the compressive stress is too high it could lead to rupture at places where there are steps in the surface of the device. It is therefore desirable to be able to measure the compressive stress in a passivation layer. This is usually done with a test sample, in the form of a small glass strip, which is coated at the same time as the silicon wafers. The sample is bowed by the mechanical stress in the passivation layer.

A method that is often used for determining the mechanical stress is to measure the radius of curvature of the bowed sample optically [1], e.g. by means of a laser interferometer. The stress can easily be cal-

[1] A. G. van Nie, A method for the determination of the stress in, and Young's modulus of silicon nitride passivation layers, Solid State Technology **23**, No. 1, 81-84, Jan. 1980.

Fig. 3. Arrangement for measuring the mechanical stress in the passivation layer of an IC. The displacement of the pins is measured by the two electronic devices on the right. The spring-pressure gauges and the stand on which the sample is set up are mounted in a frame built up from components of the Philips Variable Building System.

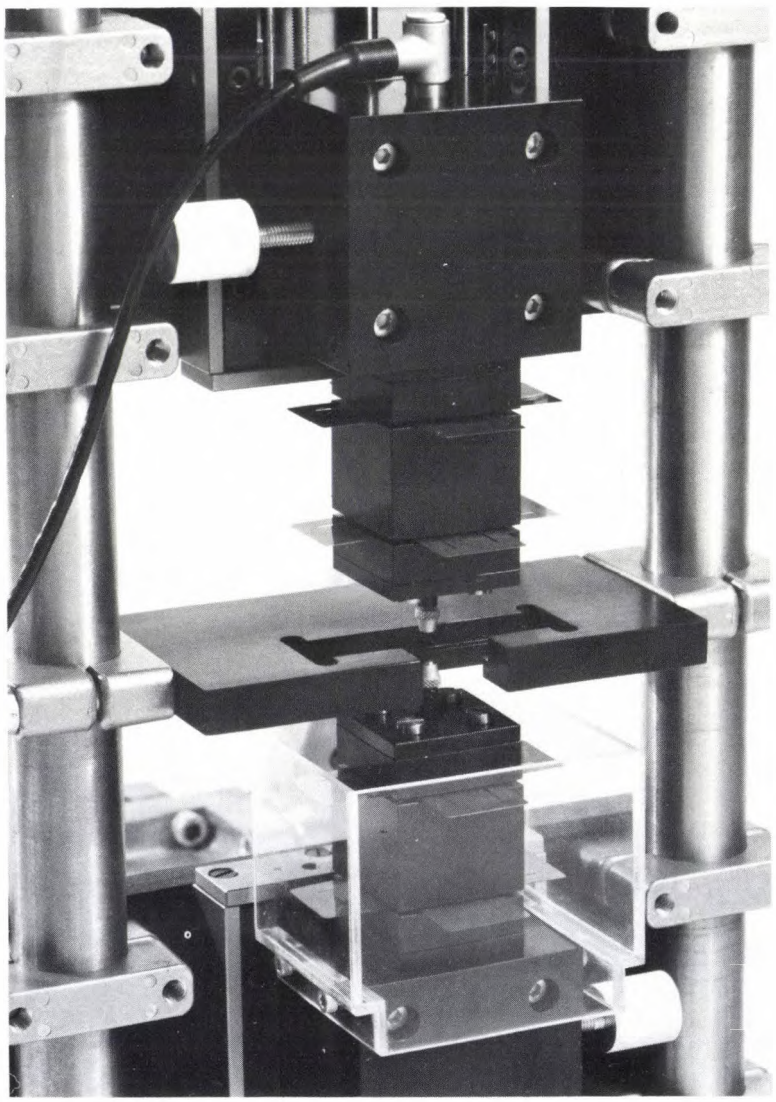


Fig. 4. Central part of the experimental arrangement showing the two spring-pressure gauges. In the lower part the Perspex protecting cap has been removed from the housing to give a clearer view of the protruding parts of the leaf springs.

culated, since it is inversely proportional to the radius of curvature.

An optical measurement of this type is very accurate, but for production purposes the method is too complicated and too expensive. We shall now describe a completely different method of determining mechanical stresses in passivation layers. This method is not quite so accurate, but is easy to use and certainly accurate enough for use in production.

In principle, the new method amounts to a determination of the force required to straighten the bowed coated sample. This force is proportional to the compressive stress to be measured. The measuring equipment is shown diagrammatically in *fig. 1*. Two spring-pressure gauges are attached to carriages (not shown) that can move in a vertical direction. The bottom end of the upper gauge and the top end of the lower gauge act as the stops. The upper spring-pressure gauge is lowered until it touches the surface of the sample, which is curved upwards. The movement is continued until the sample is straight again. Its straightness is determined by the lower gauge, which is preset to a height such that the top end is located just above the plane through the sample support points. This height is determined by a previous measurement om a small uncoated — i.e. flat — glass strip placed on the support points.

The compressive stress S in the layer when the sample is straight is given by

$$S = A(F_1 - F_2). \tag{1}$$

The force F_2 corresponds to the displacement Δx (see *fig. 2*) and is therefore known beforehand; the force F_1 is measured.

If the thickness of the samples is small compared with the other dimensions — our samples are 150 μm thick — the proportionally constant A can be calculated from Timoshenko's equations for pure bending of an infinitely thin sheet [2]. From these equations it can be shown that

$$S = \frac{E}{6(1-\nu)} \cdot \frac{t^2}{rd}, \tag{2}$$

where t is the thickness of the sample, d is the thickness of the passivation layer (about 1 Δm), E is Young's modulus of elasticity, ν Poisson's ratio, and r the radius of curvature of the bowed sample. In deriving equation (2) the effect of the passivation layer on the stiffness of the sample has been neglected, because the layer is much thinner than the sheet.

To a first approximation the shape of the strip when curved by the stress S is an arc of a circle. If the deflection y_s (the sagitta) is small with respect to the chord l, then:

$$y_s = \frac{l^2}{8} \cdot \frac{1}{r}. \tag{3}$$

From equations (2) and (3):

$$y_s = \frac{3}{4} \cdot S \cdot \frac{1-\nu}{E} \cdot \frac{l^2 d}{t^2}. \tag{4}$$

The deflection caused by a force F applied to the centre of the sample can also be calculated from Timoshenko's equations. This gives the equation

$$y_s = \frac{1}{4} F \cdot \frac{1-\nu^2}{E} \cdot \frac{l^3}{bt^3}, \tag{5}$$

where b is the width of the sample. By equating the right-hand sides of (4) and (5) we obtain the relation between the applied force F and the stress S:

$$S = \frac{1}{3} F(1+\nu) \frac{l}{dbt}. \tag{6}$$

The proportionality constant A then follows from the equations (1) and (6):

$$A = \frac{1}{3}(1+\nu) \frac{l}{dbt}. \tag{7}$$

The distance l between the support points in our arrangement is 54 mm; for the samples we used, $b = 12$ mm, $t = 150$ μm, $d = 1$ μm and $\nu = 0.22$, so that A is equal to 1.22×10^{10} m^{-2}.

Fig. 3 gives a general view of the measuring equipment, and *fig. 4* shows a detail of the central part where the two spring-pressure gauges are located. The springs are cross-shaped leaf springs of phosphor

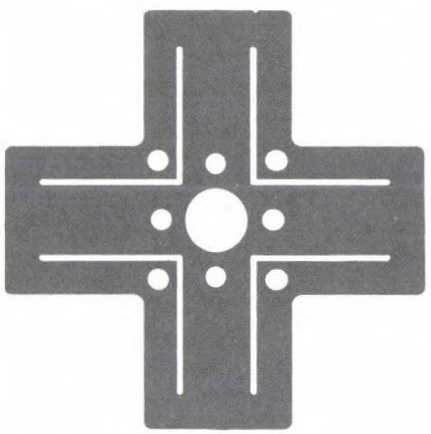


Fig. 5. The cross-shaped leaf springs (actual size). The pin passes through the central hole. The four large mounting holes are used for securing the springs to the housing, the other four are for attaching them to the pin.

Table I. Measurements on five glass test samples coated with silicon nitride under identical conditions. Three measurements were performed on each sample. The compressive stress S has been calculated from the average value $\bar{y}_s$ of the measured displacements y_s of the pin St_1 (fig. 1).

	y_s (μm)			$\bar{y}_s$	S (10^8 N/m^2)
1	33.4	33.4	33.4	33.4	3.93
2	33.8	34.0	34.0	33.9	3.99
3	33.3	33.1	33.2	33.2	3.91
4	34.4	34.3	34.4	34.4	4.05
5	32.4	32.3	32.2	32.3	3.80

[2] S. Timoshenko and J. N. Goodier, Theory of elasticity, 2nd edn, McGraw-Hill, New York 1951, pp. 225-257.

bronze; see *fig. 5*. Each gauge has two such springs, giving a good guiding action. The springs are firmly clamped to ensure that the pin always returns to the same starting point; for small deflections ($< 100\ \mu m$) the measured displacement is therefore very accurately proportional to the force. Commercially obtainable sensors are used for the pressure pins and the stop pin, with the rubber seals and guides removed.

Table I gives a summary of measurements that we made to provide an indication of the reproducibility. Five samples were all subjected to the same deposition procedures, and three measurements were made on each sample. The stop was reset after each set of measurements. It can be seen that the scatter in the measurements on any one sample is much smaller than the differences between the samples — which theoretically should be zero. The accuracy of the measurements is therefore amply sufficient for manufacturing purposes.

Summary. Since the mechanical stress in a passivation layer must lie between defined limits — if it is too low, pinholes and microcracks may occur; if it is too high there is a risk of rupture at steps in the surface — it is desirable to be able to measure it in a simple manner. The new method described here uses two spring-pressure gauges with electronic indication to measure the force required to straighten a glass test sample coated with a passivation layer. The sample becomes bowed because of the mechanical stress in the passivation layer. Each spring-pressure gauge has two cross-shaped phosphor-bronze leaf springs. The reading is accurately linear. The accuracy is amply sufficient for manufacturing purposes.

Modelling and simulation as an aid in designing a computer

J.-P. Behr, P. Pernards, B. Schendel and J. Schwandt

The behaviour and characteristics of a large system, such as a computer system, can be analysed in the design stage by making a model of the system and simulating its operation on a computer. The SAMEN/SAMO *program package is an aid to this design approach. One of the special features of the package is that the system to be simulated is formally described by means of evaluation nets ('E-nets'). This description can be used later in executing the simulation.*

Introduction: modelling and simulation

In the design of computer systems the problems encountered are similar to those found in the design of a telephone network, a traffic-control system, or a public-utility building such as a post office, to name a few examples. All these involve a 'system', that is to say a collection of parts that can each be considered as an independent entity that could function in its own right, but must also cooperate with the other parts in the system, use data from them and perhaps wait for them.

This division into independent parts may be made at different levels. In the example of the national telephone network, a division can be made into districts, or into local exchanges within a district or even, at a level of extreme refinement, into selectors, stepping switches and relays within an exchange. All these cases involve traffic flow. The traffic may consist of cars, goods, people, etc., but it may also consist of telephone calls, programs, numbers, or, in a very detailed case, electrical pulses.

In all these cases there is a flow of traffic that may be of a random nature, and there is traffic *handling*, that is to say there are a number of handling stations that although essential to the process, can be regarded as obstacles to the flow of traffic. For example, in the supermarket the situation is determined by the number of customers, the number of cash registers operating, the time it takes for the assistant to attend to each customer and the lengths of the queues.

A very important characteristic of these systems is that they are 'time discrete', i.e. their activity can be considered as a series of 'events', a series of well-defined instants at which changes occur, and not a continuous process (such as a chemical process, where materials are formed continually and other materials are used up, while at the same time the quantities change continuously). In the example of a post-office counter, the customer arrives at a particular time, the clerk appears at another *definite time* and only then can the transaction start. This transaction may take some time; at some later time the transaction ends and the letter or parcel is ready to go on its way.

Traffic problems of this kind are generally not very amenable to mathematical treatment and analysis; insight into traffic situations can only be obtained by making a model of the system and using it to simulate the actual behaviour. The model contains all the system parameters that are considered to be significant for its behaviour. In this way its is then possible to derive from the model information such as queue lengths, waiting times and 'throughput' rates.

Models at different levels of detail will sometimes be required. In town planning, for example, the post office may be regarded as a 'black box' where there are incoming and outgoing streams of people and a certain waiting time, whereas in the design of the post office itself it must be divided into doorways, stairs and corridors, counter positions and telephone boxes, etc.

Applications to computer systems

A computer system also has traffic problems of this kind. Clearly, there is an interaction between the computer *hardware* and the *software* or programs it handles. The software can be divided into user programs, and system programs, i.e. the programs that

Dipl.-Ing. J.-P. Behr, Dr Ing. P. Pernards and Dipl.-Ing., Dipl.-Wirtsch.-Ing. J. Schwandt are with Philips GmbH Forschungslaboratorium Hamburg, Hamburg, West Germany; Ing. grad. B. Schendel, formerly with these laboratories, is now with Valvo Röhren- und Halbleiterwerke der Philips GmbH, Hamburg.

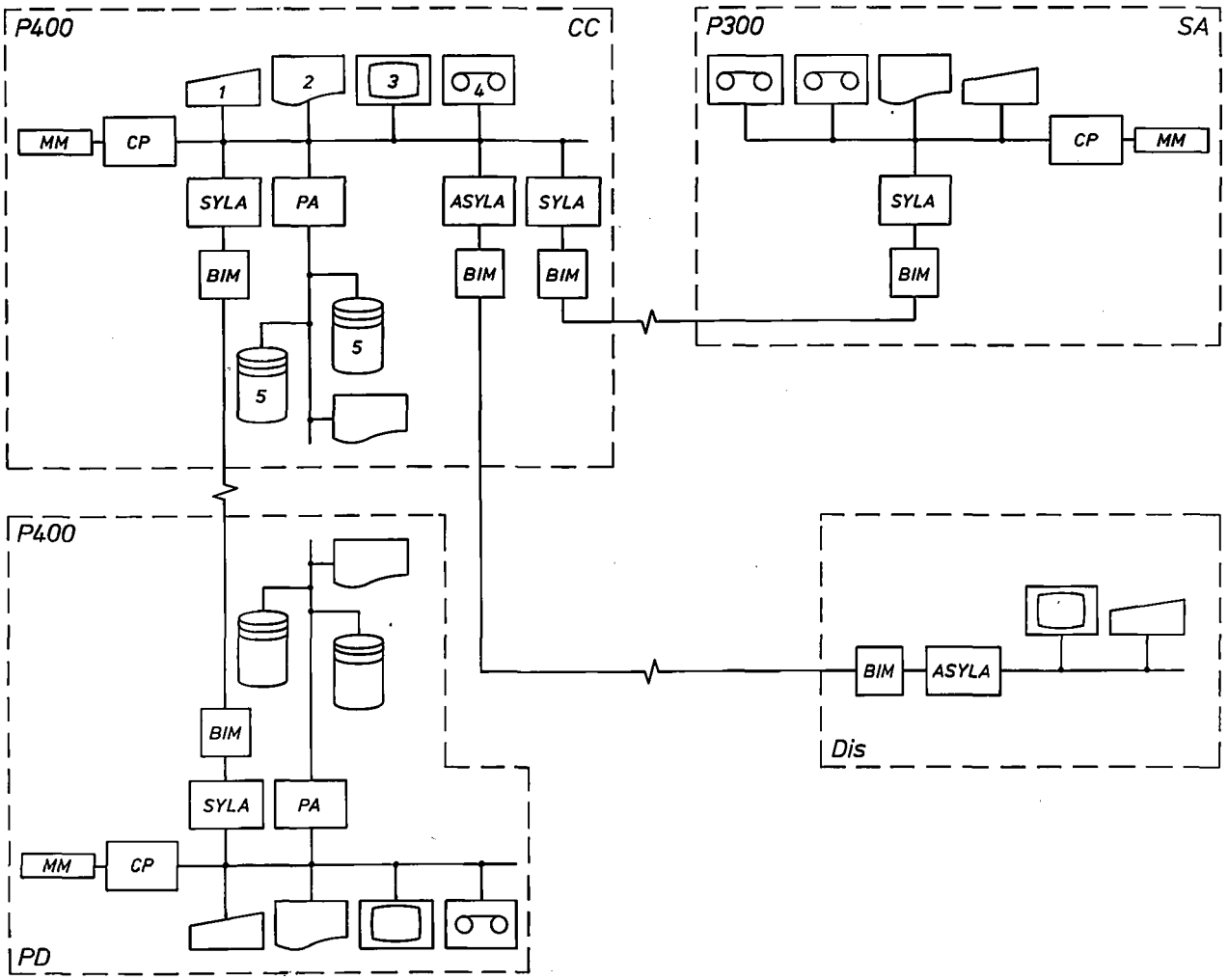


Fig. 1. Factory inventory-control system. This diagram gives an impression of the complexity that may be present in even a relatively small system. It consists of a computer centre *CC*, with a P400 computer, and three stations: one in the purchasing department *PD*, also with a P400, one in the store administration *SA* with a P300, and a display unit *Dis* in the store. *1* keyboard. *2* printer. *3* cathode-ray-tube display. *4* cassette-memory unit. *5* disc-memory unit. *CP* central processor. *MM* main memory. The asynchronous and synchronous line adapters *ASYLA* and *SYLA* and the built-in modulator-demodulator *BIM* are necessary for information traffic over longer distances. The peripheral adapter *PA* organizes traffic over short distances.

are in the system permanently for processing the user programs. The hardware can be divided into the central processor, or possibly a number of processors, the various kinds of stores or memories, the various terminals and the channels linking them all together.

Fig. 1 shows a factory inventory-control system as an example of the complexity that may be found in even a relatively small system. It is a system with terminals and consists of a P400 central computer with three tributary stations, one of which is another P400 computer, one a P300 computer and the other a display station. Each of these stations has the appropriate peripherals, such as control desks, tape and disc memories, printers, etc. The system also has special equipment for transmitting the signals over longer distances, such as synchronous and asynchronous line adapters. The behaviour of such a system could very well be the subject of a simulation study. The traffic inside the system may consist of complete jobs, or, in more detail, of individual data, characters or numbers, or even individual binary digits (bits), which are processed at various processing stations during various processing times. In extreme detail the system may be seen as a combination of standard electronic circuits and the traffic as the electrical pulses flowing in the circuits.

The study of computer systems by making a model and simulating their behaviour is very important both

in the design phase of the system, so that design faults can be traced and corrected at an early stage, and in the marketing phase, so that customer's wishes with regard to the configuration of his installation can be examined for feasibility and the customer can be advised accordingly. In the following sections we shall try to give the reader some idea of the programs that have so far been developed at the Philips Hamburg laboratories under the names SAMO (Simulation And MOdelling) and SAMEN (Simulation And Modelling with Evaluation Nets). The examples given have been worked out in cooperation with Philips GmbH Data Systems, Werk für Daten- und Informationstechnik, Siegen (Federal Republic of Germany).

Models and their description

There are two ways of making models of computer systems. Firstly, a new model could be made for each new case. This would probably have the advantage of giving fast programs, but it would entail a great deal of programming work. For this reason we have chosen a different approach in which the model is initially kept as general and simple as possible, but the options are kept open for greater detailing where necessary. At the same time the structure of the model should correspond as closely as possible to the structure of the system to be studied. As noted earlier, the system can be regarded as a collaboration between three sectors: the user programs, the system programs and the hardware. This leads to the design of three separate models: an 'application-software model' (ASM), a 'system-software model' (SSM) and a 'system-hardware model' (SHM), which are required to cooperate during the simulation phase. The application-software model serves as the origin of all the activities during the simulation phase. It consists of a list of jobs or activities that are representative of the problems the user will solve with his installation. An extensive and systematic analysis has been made of some major application areas and a library has been built up of jobs that are typical in these areas. The list is described in a notation independent of any system realization. A choice from this library can be made for each specific case, to serve as application-software model for the simulation.

The system-software model consists of a list of the machine instructions to be simulated together with a diagram representing the structure of the operating system.

The system-hardware model has to describe the equipment, possibly including the operator in charge of the computer system. The model should preferably have the same structure as the actual system and is therefore built up as a number of black boxes and their connections. Depending on the amount of detail it may be possible to subdivide the black boxes further, giving the greatest versatility for the design of the model.

The system-hardware model and the structure of the operating system can be described very well by the evaluation nets discussed below. However, since in the actual system all the processes of the operating system are run on the system hardware, the evaluation-net concept had to be extended by the 'ME-transition scheme', which will be explained later.

The simulation procedure

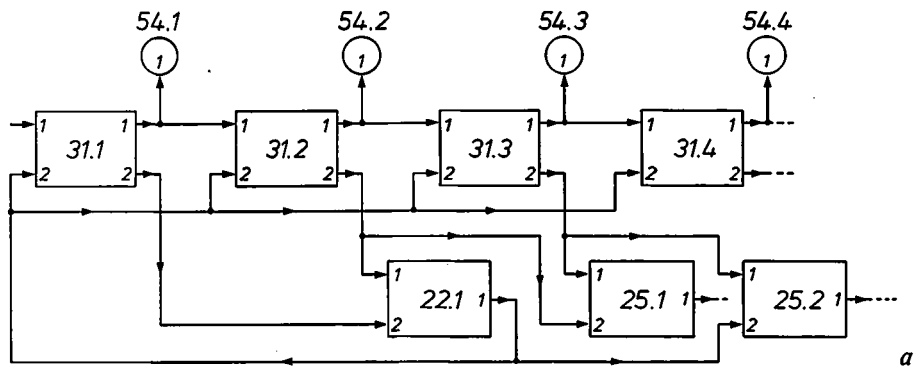
It may be evident that a model 'description' implies that a recipe can be produced from this description in the shape of a computer program for simulating the behaviour of the system on an existing computer. In event-controlled simulation the events in the simulation model are handled in the same order as they occur in the system. The collection of events is split up in a list of 'current events', events that all take place at the actual instant of model time, and a list of 'future events' that take place at a later time. An event, such as a signal change at the input of a black box, may result in an immediate change at the output — this event is then added to the list of current events — or it may result in some change taking place after a certain delay — which is then introduced as a future event and added to the list of future events.

This process of tracing the results of one event is carried out for every one of the events that occur in the list of current events. When this list has been worked through completely, the lists of future events for all the black boxes must be searched to find the event next in time and all other events that occur at the same time. They must then be arranged as a new list of current events to be worked through. The treatment of a certain event may therefore introduce other current events as well as future events. For this whole process the application-software model should in some way generate the input events. The execution of this process on a computer, i.e. searching, rearranging, administering times and events and the computer time it takes, has nothing to do with the real-time duration of the process being simulated.

The SAMEN/SAMO program package

Methods of system description have been developed at Philips Hamburg and used for the simulation of systems, resulting in the SAMEN/SAMO program package. To keep the system description clear and dis-

tinct and yet as versatile as possible, the system is divided into subsystems with their interconnections. The subsystems can be treated as individual black boxes or further subdivided as required. At the more general SAMO level of the SAMEN/SAMO program package the description of the subsystems follows a convention called BOMOL (BOx-MOdel description Language), whereas the connections are described in COMOL notation (COnnection-MOdel description Language). Both are best illustrated by an example.



cr 31.1.2 = 22.1.2 basic form

cr 31.2.2 = 22.1.1 + 25.1.2 multiple connection, expressed by the plus sign

cr 31.3.2 = 25(1 + 2)1 multiple connection; equal numbers are taken outside the brackets

cr 22.1.1 = 31(1 + + 4)2 + 25.2.2 complex multiple connection; two plus signs mean 'from ... to'

cr 31(1 —2— 3)1 = 32(2 —3— 4)1
cr 31(1 — — 4) = 54(1 — — 4)1
} repetition of similar connections, expressed by the minus sign; two minus signs mean 'from ... to'

b

Fig. 2. Model structure and its connection description. *a*) Model structure built up from five 'black boxes' types *22, 25, 31,* and *54. b*) The connection description in COMOL. (General form: *cr* ⟨output terminal denotation⟩ = ⟨input terminal denotation⟩.) The letters *cr, cb* derived from '*r*eal' or '*B*oolean' preceding the connection indicate the type of signal at the connection.

Fig. 2a shows a diagram of a network of eleven black boxes, belonging to four different types, indicated by type numbers *22, 25, 31* and *54.* Different boxes of the same type are given rank numbers, e.g. *54.1* to *54.4,* thus allowing for an arbitrary number of copies of a type that has already been described.

The basic form of the COMOL notation for a connection reads as follows:

cr ⟨output terminal denotation⟩ = ⟨input terminal denotation⟩.

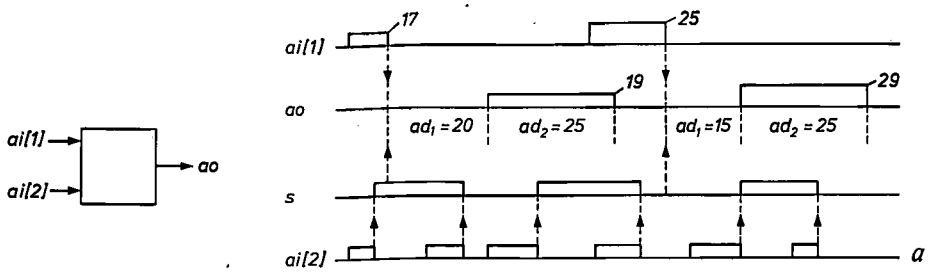
The letters *cr* indicate that this connection will be carrying signals with arbitrary values, called in most programming languages real variables, as distinct from *cb* for connections carrying Boolean (i.e. logic) variables. The sequences ⟨.....⟩ must be replaced by the actual designations for the type of box, the rank number of the box, inputs and outputs. Fig. 2*b* shows six variations of such designations, which should be self-explanatory in conjunction with fig. 2*a*.

With these conventions it is obviously easy to ensure that information appearing at an output terminal during the simulation is passed on to the appropriate input terminal.

It is not nearly so easy to describe the characteristics of the BOMOL conventions for the description of black boxes, because of the freedom with which the user can define the functions of a black box. His definitions can include the use of state diagrams, flow charts and evaluation nets, or just verbal descriptions. The method of evaluation nets will be dealt with more fully in the next section; here we shall discuss the situation given in *fig. 3.* A black box, which should be known by a type number, is shown in fig. 3*a*, with two inputs and one output. The timing diagram and the verbal description define the process that takes place inside the box. The BOMOL description is based on the use of ALGOL as a programming language; the structure of these programs as well as a number of conventions to be used, e.g. for the identifiers, are laid down as the rules of BOMOL. Thus the labelling of the inputs as *ai[1]* and *ai[2]* and of the output as *ao* is part of the BOMOL conventions, which say that BOMOL symbols with a standard significance must

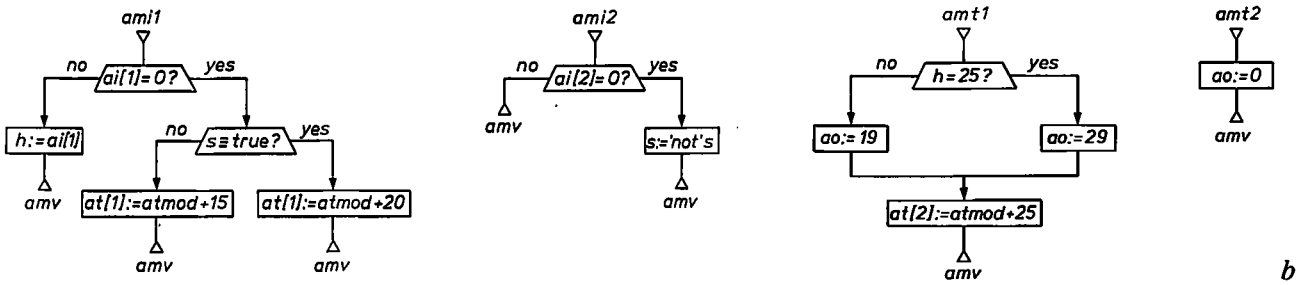
[1] K. Lagemann, Das Simulationsprogrammpaket „SAMO", Angew. Informatik 16, 488-492, 1974.

Fig. 3. Various stages in the black-box description in SAMO.



a

a) Timing diagram and verbal description. The black box is characterized by the following rules:
— 0-1-0 signal sequences are transmitted to input *ai[2]*; the 1-0 transitions effect an instantaneous change in the Boolean parameter *s*.
— Each trailing edge of the pulse sequence transmitted to the input *ai[1]* generates a pulse of length $ad_2 = 25$ time units after a delay ad_1 at the output *ao*.
— The delay ad_1 totals 20 or 15 time units, depending on whether the state variable *s* has the value 1 or 0 when the delay starts.
— The input amplitude at *ai[1]* can have either of the values 17 or 25 selected at random. The corresponding output amplitude becomes 19 or 29 respectively.



b

b) The flow chart.

```
        'begin'
        'own"integer' h;
        'own"boolean' s;
ami1:   'if' ai[1] ≠ 0 'then"begin' h:= ai[1]; 'goto' amv; 'end';
        at[1]:=atmod + ('if' s 'then' 20 'else' 15); 'goto' amv;
ami2:   'if' ai[2] = 0 'then' s:='not' s; 'goto' amv;
amt1:   ao := 'if' h=25 'then' 29 'else' 19; at [2] := atmod + 25; 'goto' amv;
amt2:   ao := 0; 'goto' amv;
        'end';
```

c

c) Description in BOMOL (BOx MOdel description Language). Characteristics of BOMOL are:
— The programming language is ALGOL.
— All changes at the inputs are represented by means of entry-point labels *ami*; thus signal changes at input *ai[2]* are represented by entry-point label *ami2:* in the case of a trailing edge the value of *s* changes. Similarly for changes at input *ai[1]* entry-point label *ami1* is used: after a leading edge (i.e. $ai[1] \neq 0$) the signal value is stored in an auxiliary variable *h*. After a trailing edge the delay time $ad_1 = 20$ or $ad_1 = 15$ is set to a value depending on the value of *s*.
— Delayed actions inside a black box are represented by entry-point labels *amt;* thus after a delay ad_1 operations are resumed via entry-point label *amt1;* the start for delay $ad_2 = 25$ is initiated here; the next action, resetting the output signal $ao = 0$, takes place via entry-point label *amt2*.
— All predetermined identifiers start with the letter *a*. Arbitrarily chosen identifiers such as *h* and *s* must be declared at the beginning of the program.
— Exits are represented in the program by 'goto *amv*'.

start with the letter *a*. In fig. 3*b* the model description is translated into a flow chart. A BOMOL convention for reading this chart is that every event taking place at an input *ai[1]* (e.g. a signal change at *ai[1]*) gives access to the flow chart at label *ami1*, and accordingly for input *ai[2]* at label *ami2*. When the former happens because *ai[1]* becomes equal to 0 at a time recorded in the variable *atmod* (model time), a future event (the start of the output signal) will be produced after a delay ad_1 and an 'alarm clock' is set

to $atmod+ad_1$ to record this future event. The flow chart shows that, in accordance with the description, the value of ad_1 depends on the value of s.

BOMOL allows a set of several of these alarms to be arranged in a clock array: *at[1]*, *at[2]*, etc. When the moment for treating the future event *at[1]* has arrived, the program is entered via label *amt1*, and accordingly through label *amt2* for the time recorded as *at[2]*. After dealing with the event in question, the program leaves the treatment of the particular box, to treat further boxes occurring in the list of current events for this same instant *atmod*. The way in which this is organized cannot be discussed here: it takes place in a supervisory area of the SAMO simulation-program package.

From the diagrams in fig. 3*b* it is now easy to write the actual program as an ALGOL program. Other variables the user requires, apart from those with a standard BOMOL significance, such as *h* and *s* in our case, have to be declared at the start of the program in the normal way, e.g. 'own integer *h*' and 'own Boolean *s*' (see fig. 3*c*).

Evaluation nets

In order to offer the user additional support in model description, evaluation nets were chosen as a formal description method, and a 'SAMEN level' was added to the simulation-program package. The SAMEN description of the black boxes, based on evaluation nets (E-nets), can replace the BOMOL part of the simulation program. E-nets, introduced by G. J. Nutt [2], as an extension of Petri nets [3], have proved to be the most satisfactory of all methods for describing time-discrete processes [4,5]. We shall explain the E-nets from the example of a post-office counter position. The situation is characterized by the fact that a clerk and a customer are both required for the transaction.

In the pre-serving situation only one of them is available. As soon as both are present we have a 'transition' to the serving situation, which may last for some time. At the end of this situation, we have a transition to the post-serving situation, where the clerk is free to help a new customer and the earlier customer is free to leave the post office. This second transition would involve no handling time.

The E-net for this example is shown in *fig. 4*. The two transitions are indicated by the vertical bars *a1* and *a2*. The two requirements that characterize the pre-serving situation, the clerk and the customer, are shown as 'locations' *b1* and *b2*; the serving situation is indicated by one location *b3*, and the free customer by the location *b4*. The locations show the potential requirements; when they are fulfilled this is shown by placing a 'token' in the relevant location. A transition is said to 'fire' — e.g. the transaction starts — as soon as the input locations have received a token and the output location has become empty. After a transition time (if necessary), tokens from the input side move over to the output side, leaving the input locations empty. In this way a flow of tokens through the E-net represents all the transactions in the system.

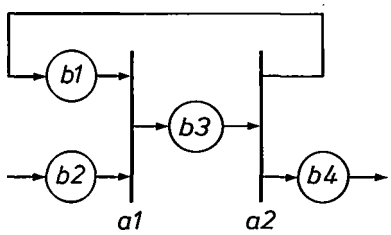


Fig. 4. Evaluation-net description of a post-office counter position. *a1* and *a2* represent 'transitions': *a1* indicates the start of the transaction and *a2* its end. *b1* to *b4* represent 'locations', i.e. potential requirements. *b1* clerk available. *b2* customer has entered post office. *b3* clerk serves customer. *b4* customer ready to leave post office. When a requirement is fulfilled, the location carries a 'token'.

Tokens may carry with them, in the form of 'attributes', information required in the process; e.g. the customer token might carry as an attribute the handling time for the customer's particular problem. In this case the transition *a1* would have to look up this attribute and adjust the transition time accordingly. Also, attributes might change their value in the course of the transition.

Fig. 5 shows the five basic types into which transitions are usually subdivided. In addition to the 'transition procedure' that determines how the attributes are modified and the 'transition-time procedure' that determines the transition time, the *X*- and *Y*-transitions also require a 'resolution procedure' to determine the routing of the tokens (indicated by '*r*' in fig. 5). While working with E-nets it was found that two additional types of transition were required [6]: the *TI*-transition, for indicating transitions that can be interrupted and terminated to simulate interrupts and failures in computer systems (*fig. 6*), and the

[2] G. J. Nutt, Evaluation nets for computer system performance analysis, AFIPS Conf. Proc. 41, 279-286, 1972.
[3] C. A. Petri, Kommunikation mit Automaten, Thesis, Bonn 1962.
[4] J.-P. Behr, R. Isernhagen, P. Pernards and L. Stewen, Modellbeschreibung mit Auswertungsnetzen, Angew. Informatik 17, 375-382, 1975.
[5] J.-P. Behr, R. Isernhagen, P. Pernards and L. Stewen, Erfahrungen mit Auswertungsnetzen — Implementierung, Alternativen, Angew. Informatik 17, 427-432, 1975.
[6] L. Stewen, Auswertungsnetze als Hilfsmittel zur Modellbildung — Probleme und deren Lösungen, Lecture Notes in Computer Science 34, 462-474, 1975.

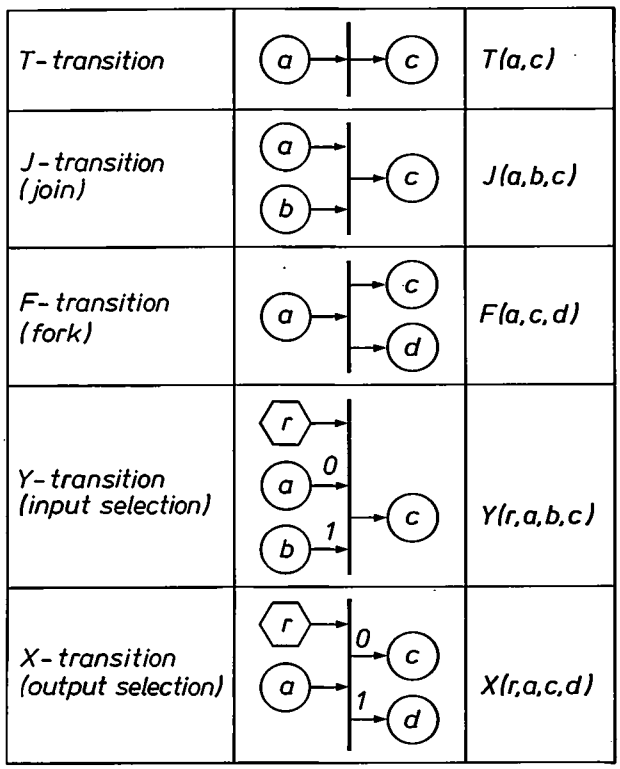


Fig. 5. Basic transition types in evaluation nets. A transition is said to 'fire' if the input locations contain a token (i.e. if the input conditions are fulfilled) and the output locations are empty. The firing may take some specified transition time (delay). As soon as this has elapsed tokens from the input side move over to the output side. The *J*-transition thus requires two tokens in *a* and *b* to produce one in *c* after firing. For the *F*-transition, one token in *a* after firing produces tokens in *c* and *d*. The *Y*-transition may describe a conflict case: when *a* and *b* are both full at the instant when *c* becomes empty, priority is controlled by a resolution procedure given by *r*. On the other hand if only *a* or *b* contains a token while *c* is empty the token is moved without calling the resolution procedure. Such a procedure in the *X*-transition determines the routing of tokens to *c* or *d*.

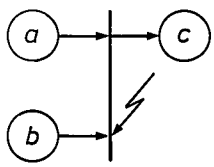


Fig. 6. *TI*-transition. The *TI*-transition is a *T*-transition with a special interrupt input location *b*. It behaves like a *T*-transition and 'fires' in the normal way if *b* does not contain a token, i.e. after a specified delay, the token in *a* moves on to *c*. If *b* receives a token while the *TI*-transition is in the delay phase (interrupt condition) it fires immediately, regardless of the remaining transition time. A token arriving at *b* has no effect if *a* is empty.

ME-transition to indicate the change from a low-detail to a high-detail level of description (*fig. 7*).

Fig. 8 shows a simple example of an E-net for a central processor *CP* with two peripherals or terminals, *T1* and *T2*. Jobs come in from the terminals into locations *b1* and *b2*. As these locations belong to both *CP* and terminal, they are shown as semicircles. Tokens in this example have two attributes: *K(1)*, the number of the terminal of origin, and *K(2)* the processing time for the particular job.

Tokens in the E-net are not in fact usually shown; their positions would only allow a rendering of an instantaneous situation. Keeping track of these movements is the task of the simulation program. It does this by numbering tokens as they enter the system and keeping a record of which token is in which location.

If *b1* and *b2* should simultaneously receive a token while *b3* is empty, the resolution rule *r1* determines which of the two may go through first. Transition *a2* refers to taking over the job into the main memory, which can only be done if the central processor is not engaged (indicated by *b5* empty and *b4* occupied). Transition *a3* refers to processing by the central processor, with the time required taken from the second attribute of the token. When the processing has been completed and the token arrives in *b6*, the resolution rule *r4* determines which terminal the results should return to from the first attribute of the token.

The SAMEN input conventions

E-nets not only have the advantage that they provide the user with a clear insight, in the form of a drawing, into the structure of his problem; they can also be clearly described in a formal manner. This is essential when producing the simulation program [5]. As an example *fig. 9* shows the transition *a4* of fig. 8. This is an *X*-transition with locations *b6*, *b7* and *b8* and a resolution rule *r4*. Since the action is only a choice, there is no transition time involved for both routings and no transition procedure is necessary for changing the attributes. Fig. *9a* shows the way this is described formally as defined by Nutt. The part *S* gives the structural information, *T* the time information and *Q* the transition procedure. The notation *b6[2]* indicates that the tokens have two attributes, which are denoted by *b6(1)* and *b6(2)* in *r4*. *M(x)* means 'value of *x*'. The notation used here closely resembles the ALGOL notation. It is therefore easy to change it into the SAMEN description given in fig. *9b*. (Symbols with a standard significance in SAMEN start with the letter *b*.) Here BXPTRA is the name of a procedure written in ALGOL that handles the *X*-transitions in accordance with the given parameters. This procedure, like the analogous procedures for all the other basic transitions, is a standard ALGOL program for SAMEN, and in fact together with the SAMEN-control routine these procedures constitute the SAMEN 'language'.

In the example BXPTRA (.....), *Q0* is the name of an 'empty procedure' that produces no results. SAMEN also arranges the tokens, which are numbered on entry, in an array of locations such that *B⟨i⟩* represents the number of the token residing in loca-

tion *i* at that moment and *B⟨i⟩(1)* its first attribute. Thus, *B6(1)* denotes the first attribute of the token in location *6*. In this way all the transitions inside a black box are described. As the input locations of one transition are output locations for another, the interconnections are automatically established. It is thus possible to establish the E-net description in SAMEN for each of the black boxes separately: *CP, T1* and *T2*. From this the SAMEN translator can make a complete ALGOL program to carry out the simulation.

Interactive model and simulation studies

A very important aspect of the use of computers as an aid in the design of systems is the interface between the user and the computer installation. The provision of terminals in the designer's office and dialogue facilities with the host computer system are a great help here. These considerations led to the writing of a set of programs called IMOBS (Interactive MOdel Building and Simulation), which is arranged in two program packages.

The package GIPSIE (Graphical Interactive Program for SImulation with Evaluation nets) can be used to build up the diagram of an evaluation net on a graphic-display terminal. Transitions, input and output locations and their connections can be displayed at any desired position on the screen. Instructions can be keyed in to shift parts of the figure or remove them. An example is given in *fig. 10*, showing how the user of IMOBS can obtain the model of fig. 8 on the screen. If the result is satisfactory, the GIPSIE program can derive the data for the structure of the net from the graph on the display screen and generate the appropriate program codes in accordance with the SAMEN conventions. The second program package, called DIMPLE (Dialogue IMPLEmentation package) can be used in a dialogue with the computer to enter values of the model parameters and prepare the simulation.

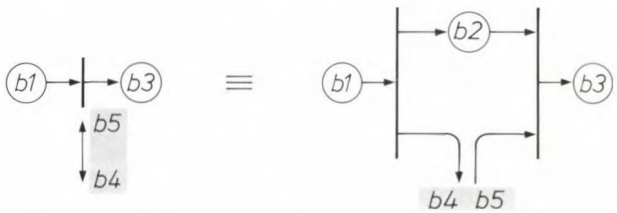


Fig. 7. *ME*-transition. This transition is useful in hierarchically structured systems i.e. systems described at different levels of detail. In the symbol on the left the double-headed arrow refers to locations *b4*, *b5*, at a description level of greater detail, and the actions in question are either taken into account in the transition function of the transition on the left, or the actions are followed in detail at the lower level, as indicated in the equivalent diagram on the right. An example of such a situation might be found in the model for an operating system, where every transition corresponds to different, possibly complicated, activities of the hardware.

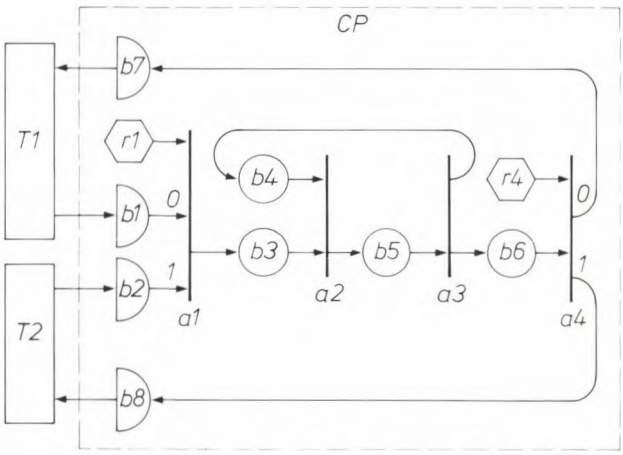


Fig. 8. Simplified model of a central processor with two peripherals or terminals. Only the E-net for the central processor is given. Locations that are common to processor and terminal are shown as semicircles. The locations have the following significance: *b1* job from terminal *1*. *b2* job from terminal *2*. *b3* job requesting *CP*. *b4* *CP* is idle. *b5* *CP* is busy. *b6* *CP* has finished job. *b7* job from *T1* is complete. *b8* job from *T2* is complete. The resolution procedures: *r1* chooses job from *T1* or *T2*; *r4* routes job to *b7* or *b8*. The job tokens have two attributes: *K(2)* to represent processing time, and *K(1)* to represent the terminal number. The transition time of transition *a3* is therefore taken from *K(2)* and the resolution procedure *r4* will inspect *K(1)* to determine the routing.

Presentation of results

Every user of model-simulation techniques will have his own ideas about the presentation of the results. In testing an actual system he will select very

```
a4 = (x(r4,b6[2],b7[2],b8[2]), (0,0), — )
       S                        T     Q

a   r4: [(M(b6(1)) = 1) → M(r4) := 0;
         (M(b6(1)) = 2) → M(r4) := 1]

    'INTEGER''PROCEDURE' R4;
    R4 := 'IF' B6(1) = 1 'THEN' 0 'ELSE' 1;
b   A4: BXPTRA (6,7,8, 0,0, Q0, R4);
```

Fig. 9. Example of the coding of a transition. *a*) Formal description of the transition *a4* in fig. 8 as initially defined by G. J. Nutt. Section *S* defines the transition itself, *T* specifies the transition time and *Q* the transition procedure. *b6[2]* indicates that a token in location *b6* has two attributes. In the description of the resolution procedure *r4*, *M(b6(1))* denotes the value of the first attribute of the token in location *b6*. *b*) Coding of the same transition in SAMEN. Here it should be known that the tokens are numbered on entry and that in SAMEN all the tokens of a black box have the same number of attributes (here two) so that they can be organized as a matrix. The number of the token residing in location *i* is stored under *B⟨1⟩*. Thus *B6(1)* denotes the first attribute of the token at location *6*. In SAMEN BXPTRA(. . . .) is a standard procedure representing the *X*-transition. Its parameters have the following significance: *6, 7, 8*: input and output locations. *0, 0*: transition times for both directions, *Q0*: transition procedure (*Q0* happens to be an 'empty' procedure, i.e. one that has no effect). *R4*: resolution procedure. The symbol := should be read as 'becomes' or 'gets the value'.

[7] J.-P. Behr, GIPSIE, ein Programm zur interaktiven grafischen Modellbildung mit Auswertungsnetzen, Angew. Informatik **20**, 60-65, 1978.

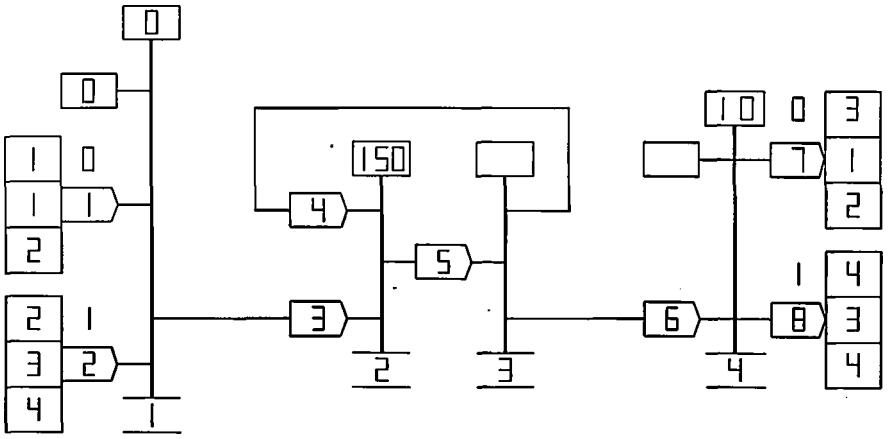


Fig. 10. E-net for the example of fig. 8 as produced with GIPSIE. This print of the information on the screen shows that some of the symbols differ from the original form. To save memory space and computer time, locations are not represented by circles but by a kind of arrow. The transition number is shown in a rectangular 'frame' underneath the transition; this frame is open on both sides (as in the example here for all the transitions) if there is no transition procedure (the procedure is 'empty' and denoted by *Q0*); otherwise the frame is closed. A rectangle above the transition shows the transition time if this is a constant; if this rectangle is empty, the transition time is evaluated by a transition-time procedure. A constant value in a resolution location characterizes a fixed priority sequence, starting at the given value. The input and output locations contain extra information required for the SAMO connection net.

specific points for measurement and a suitable measuring instrument for each point. This facility should also be available to the user of a simulation model. In BOMOL he can do this by introducing every measuring instrument required as a black-box model and connecting it to any desired point by means of the connection network in as many ways as he wishes. A library of models of a number of standard instruments, e.g. a multibeam oscilloscope (see *fig. 11*), is already available in the form of BOMOL programs and a BOMOL description can be made if required for any special instrument. It is even possible to create measuring instruments that do not in fact exist but would be useful for the application, e.g. instruments in which measured results are processed immediately. Results of various simulations can also be stored provisionally in the memory for subsequent comparison and processing, for example to evaluate them statistically, to draw histograms or to make graphs, so that in the end a diagram can be produced directly as illustrated in *fig. 12*.

Model time	Real inputs 1	2	3	Boolean input 1	Model time
0	:	:	>>>+	:	0
1	:	>>>-	: |	:	1
2	>>>+	: |	: |	:	2
3	: |	: |	<+<*	:	3
4	: |	<<<*	: |	>>+	4
5	>>>*+	:	: |	: |	5
6	: |	:	: |	: |	6
7	: |	>>>>>+	: |	: |	7
8	: |	: |	: |	<<*	8
9	<<<<*	: |	>*>>>+	:	9
10	:	>>>>>+	: |	:	10
11	>>>-	: |	: |	:	11
12	: |	: |	: |	>>+	12
13	: |	<<<<<*	: |	: |	13
14	<<<*	:	: |	: |	14
15	:	>>>+	: |	: |	15
16	:	: |	<<<<<*	: |	16
17	>>>>>>+	: |	:	: |	17
18	: |	: |	:	: |	18
19	: |	: |	:	: |	19
20	<<<<<<*	<<<*	>>>+	<<*	20
21	:	>>>-	: |	:	21
22	>>>+	: |	: |	:	22
23	: |	: |	<+<*	:	23
24	: |	<<<*	: |	>>+	24
25	>>>*+	:	: |	: |	25
26	: |	:	: |	: |	26
27	: |	>>>>>+	: |	: |	27
28	: |	: |	: |	<<*	28
29	<<<<*	: |	>*>>>+	:	29
30	:	>>>>>+	: |	:	30
31	>>>-	: |	: |	:	31
32	: |	: |	: |	>>+	32
33	: |	<<<<<*	: |	: |	33
34	<<<*	:	: |	: |	34
35	:	>>>+	: |	: |	35
36	:	: |	<<<<<*	: |	36
37	>>>>>>+	: |	:	: |	37

Fig. 11. Simulation results in the form of a pulse diagram. When simulating at a very detailed level, e.g. at a circuit level, a pulse diagram is a common method for presenting results. SAMO therefore includes a standard black box for presenting input and output values as a pulse pattern.

Study of a terminal system

As a final example we shall give a rather more detailed discussion of a simplified model of a terminal system, consisting of four terminals and a central processor, as shown in *fig. 13*. In the example, which is an extension of the one of fig. 8, both kinds of description (SAMO and SAMEN) will be used. However, we shall first describe the system and how it presents itself to the user.

Let us start to examine the problem at the moment when a user is thinking about the formulation of a

new job. During that period the terminal remains quiet. The time spent thinking in this model is obtained from a random process called 'ranerl' in the program. When the user has finished thinking, he can start to type in his job at the terminal, provided the central processor has finished any previous job started at this same terminal. If not, the user has to wait until the central processor indicates that the previous job has been finished. A list of waiting times or perhaps their distribution could be one of the results from the simulation. In this model a fixed period of time is taken for typing. At the end of this period a request is sent to the processor and the input data, which was temporarily stored at the terminal, is sent *'en bloc'* to the processor, provided the buffer set aside for buffering input data is free to accept the information. As well as feeding the problem into the buffer, additional information has to be read from the disc memory. In this model a fixed time period is allowed for this task as well. When the central processor has finished the previous calculation, the processor transfers the new problem information from the buffer into the main working memory and the processing of the new job can start. Once again it should be noted that the simulation of the behaviour of an installation has nothing to do with the actual execution of computations, because this is not necessary to obtain the results mentioned above. Simulation only covers the administration, registration and recording of traffic flowing inside the system. The signals treated during the simulation are all of the kind: 'user is thinking', 'typing starts', 'typing ends' and so on. Durations are either taken as constant or taken from statistical information derived from actual practice.

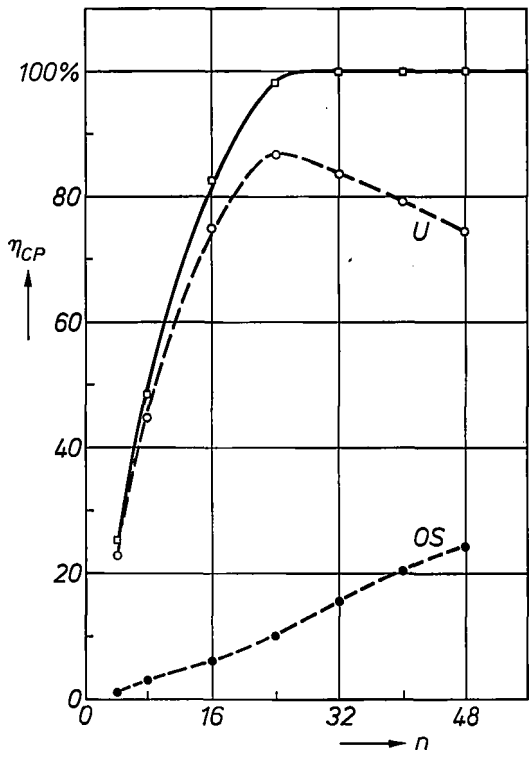


Fig. 12. Simulation results as a plot. The central-processor utilization η_{CP} is plotted against the number of terminals. The solid curve represents the total utilization, curve *U* the contribution from the users and curve *OS* represents the contribution from the operating system.

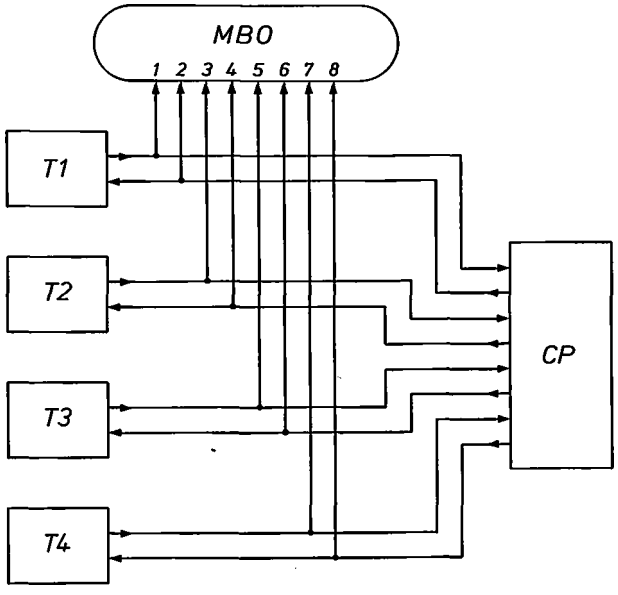


Fig. 13. Block diagram of a system with one central processor, four terminals *T1* to *T4* and a measuring instrument (e.g. a multiple-beam oscilloscope *MBO*) for displaying the signals flowing between them).

In this example the application-software model (ASM) and the system-software model (SSM) are reduced to the minimum, and at first glance it seems that only the hardware model remains. The ASM is to some extent represented in the random process for determining a thinking time. The SSM is less clearly distinguished; it would affect the job-processing time, which is now generated statistically. It could be said to be present in this random process ('rapois' in the block diagram) inside the terminal.

Diagram I illustrates the traffic as it is during the first part of the simulation run. This event list gives an impression of the kind of situations that may be expected in actual practice.

The model description at the SAMO level

The model is described in two types of black box, box *1* for the processor and box *2* for the terminals. Box *2* is thus present in four copies. Box *2* has one input *ai* for signals coming from the processor, and one output *ao* for signals going to the processor. The central-processor box has four inputs, *ai[1]* to *ai[4]*, connected to the outputs of the terminals, and four

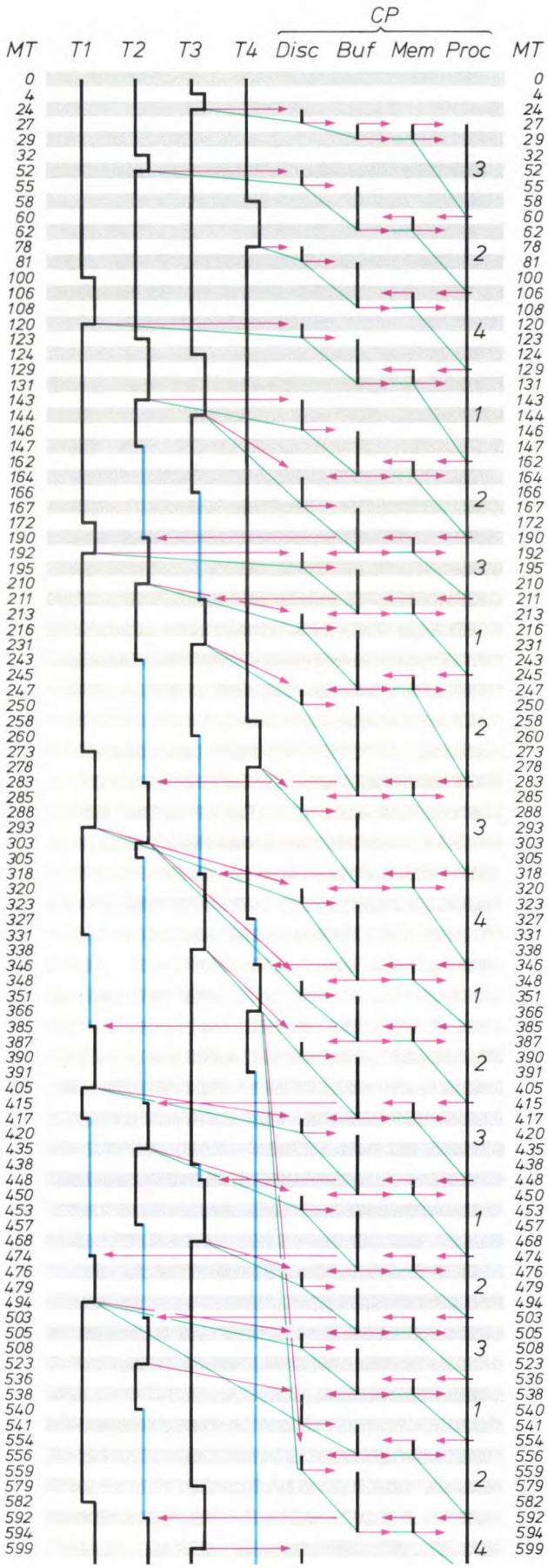


Diagram I. Event list of part of the simulation of the behaviour of the terminal system of fig. 13. The model time *MT* is shown at the left and right. The left-hand half of the diagram shows the periods the user spends at the terminals *1* to *4* in thinking about a problem (black line in the *left-hand* position), the periods when he is typing the problem in (black line in the *right-hand* position) and the periods when he has to wait before he can start to type (*blue line*). The right-hand half of the diagram shows the periods in which the central processor is reading from the disc memory *Disc*, the periods when the buffer *Buf* is full and the periods when the main memory *Mem* is being loaded. The column *Proc* indicates the activity of the central processor; the numbers *1* to *4* show which terminal the central processor is working for. The red arrows show how the events trigger other events. The thin green lines are only intended to show the 'history' of a job more easily.

The simulation shows that users soon have to wait for typing in because previous jobs at that terminal have not been finished. Reading from the disc memory first starts as soon as a terminal indicates that a new job has been typed in, but later it can only start when the buffer is empty again. It might have been more logical to show the buffer as 'full' from the moment when reading from disc starts, in the same way as it is recorded as 'empty' when loading the main memory has ended. This choice has been based on the fact that the SAMO flow charts and results only relate to 'events' and the recording is not continuous. (Note that equal times *do not* correspond to equal line lengths.) At model time *417* it happens for the first time that when the buffer becomes free, terminal *T1* and terminal *T4* both have a job ready, and the priority rule comes into force. This happens again on several other occasions and the treatment of terminal *T4* is therefore postponed until model time *556*.

outputs, *ao*[*1*] to *ao*[*4*], connected to the inputs of the terminals. As already explained, the programs are written as ALGOL programs with the BOMOL conventions. The program structure is given in the flow charts *fig. 14* and *fig. 15*. Every signal change at an input causes the program to enter at label '*ami*'. The terminal box employs two 'alarm clocks' (see p. 138), one set to mark the end of the thinking period and recorded as *at*[*1*], the other to mark the end of the typing period and recorded as *at*[*2*]. A fixed period of 20 units of time is allowed for the typing. When these events occur, the program is entered at *amt1* or *amt2* as appropriate. The method of starting the program, e.g. by calling up the procedure 'think' is not shown (in the actual program these actions are taken in the initialization phase by entering through the labels '*amnr*', '*amse*', and '*amop*'). The processor box is slightly more complicated; three alarm clocks are required here. The end of reading from the disc memory is given by *at*[*1*]; a fixed duration of 3 units is allowed for this operation. The end of loading the main memory is marked by *at*[*2*]; a fixed period of 2 units is used here. The end of processing for the job is given by *at*[*3*]; the duration is determined at the terminal where the job was initiated by means of the random process 'rapois', and handed over to the processor as the value of the output signal from the terminal. This signal therefore has a double function: it signals the event 'typing finished' and carries the information about 'processing time'. The processor inputs 1 to 4 are strobed for signals in such a way that the lower

terminal numbers have a higher priority than the others. More detail will be found in *fig. 16*, which gives the full BOMOL descriptions, known as 'Macro BQ1' and 'Macro BQ2'.

In fig. 13 the introduction of the instrument 'multiple-beam oscilloscope', whose standard description is available in the SAMEN/SAMO program library, was included to demonstrate that it can be done and to obtain a simplified version of the diagram I from the other results of the simulation run. In the COMOL description of the connection network, which is shown in *fig. 17*, this instrument has been left out.

The model description at the SAMEN level

The description of the complete example in terms of evaluation nets takes on the shape of the diagrams shown in *fig. 18*, which were generated by means of the program package GIPSIE. The processor black box is only slightly different from the one given in fig. 8. Four transitions can be distinguished: first, the message from the terminal indicating that a new job is arriving and the read-out of additional information from the disc memory with a fixed transition time of 3 units. Second, loading the job into the main memory under the control of the processor: here the fixed

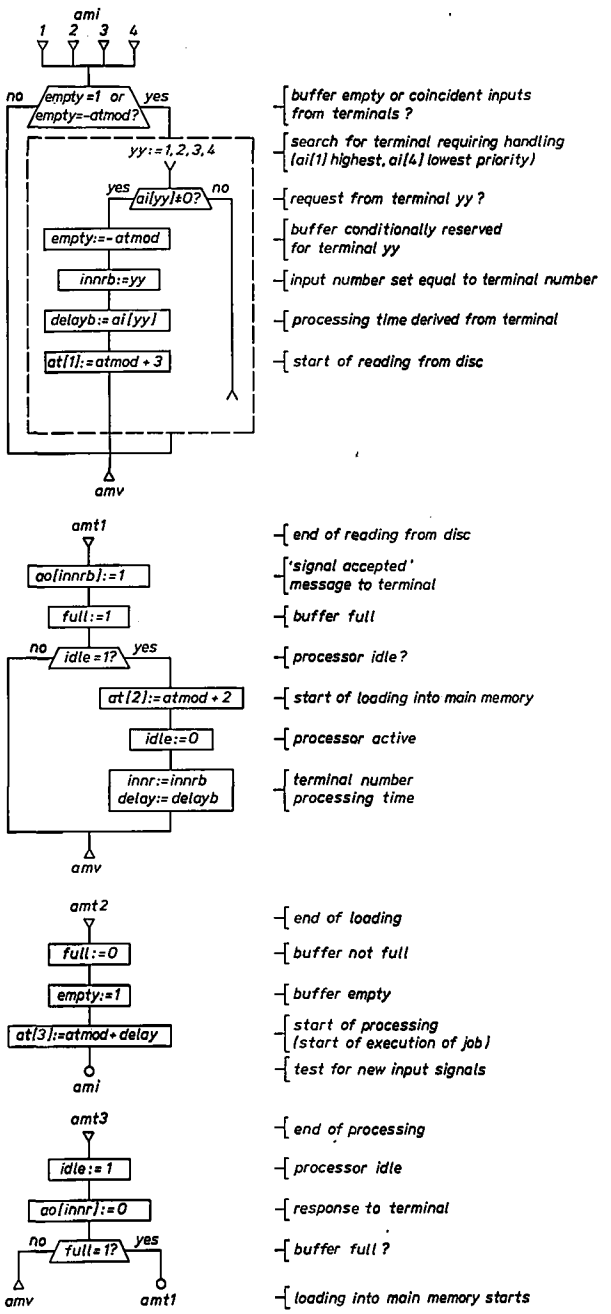


Fig. 14. Flow charts for the most important part of the central processor (box *BQ1*).

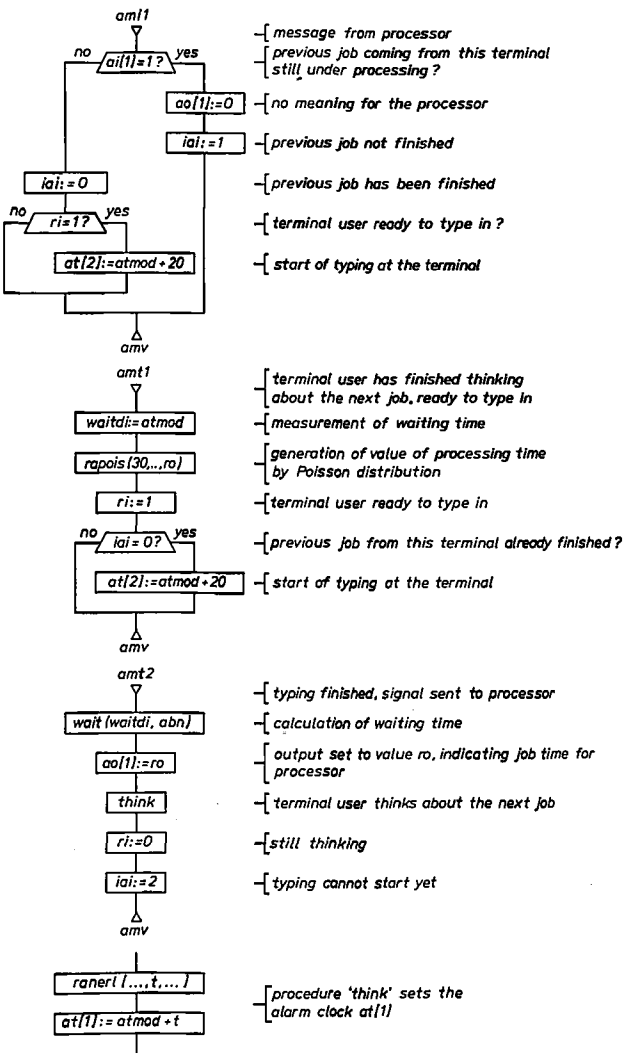


Fig. 15. Flow charts for the most important part of the terminal (box *BQ2*).

```
100.$MACRO BQ1;
200.'BEGIN'
300.'COMMENT' PROCESSOR CONNECTED TO 4 TERMINALS VIA AI(/1...4/),
400.          AI(/1/) WITH HIGHEST, AI(/4/) WITH LOWEST PRIORITY,
500.          AT(/1/): BUFFERING ON FLOPPY DISK,
600.          AT(/2/): COMPILING, LINKING, LOADING,
700.          AT(/3/): PROCESSING;
800.'BEGIN'
900.'OWN''INTEGER' INNR,INNRB,IDLE,EMPTY,FULL,DELAY,DELAYB;
1000.'INTEGER' XX,YY;
1100.
1200.'SWITCH' AMSTART:=AMNR,AMSE,AMOP,AMT1,AMT2,AMT3,AMI,AMI,AMI,AMI;
1300.'GOTO' AMSTART(/AJX/);
1400.
1500.AMNR:  ATZ:=3; AIZ:=4; AOZ:=4; 'GOTO' AMV;
1600.AMSE:  IDLE:=EMPTY:=1; FULL:=0; AFEHL:=0; 'GOTO' AMV;
1700.
1800.AMI:   'IF' EMPTY=1 'OR' EMPTY=-ATMOD
1900.       'THEN' 'BEGIN'
2000.       'FOR' YY:=1,2,3,4 'DO'
2100.       'IF' AI(/YY/)#0
2200.       'THEN' 'BEGIN'
2300.       EMPTY:=-ATMOD;
2400.       INNRB:=YY; DELAYB:=AI(/YY/);
2500.       AT(/1/):=ATMOD+3;
2600.      'COMMENT' BEGIN OF READING FROM DISK;
2700.      'GOTO' AMV;
2800.      'END' THEN;
2900.      'END' THEN;
3000.      'GOTO' AMV;
3100.
3200.AMT1: 'COMMENT' END OF READING FROM DISK;
3300.      AO(/INNRB/):=1;
3400.      FULL:=1;
3500.      'IF' IDLE=1
3600.      'THEN' 'BEGIN'
3700.      AT(/2/):=ATMOD+2;
3800.      'COMMENT' BEGIN OF LOADING;
3900.      IDLE:=0;
4000.      INNR:=INNRB;  DELAY:=DELAYB;
4100.      'END' THEN;
4200.      'GOTO' AMV;
4300.
4400.AMT2: 'COMMENT' END OF LOADING, BEGIN OF PROCESSING;
4500.      FULL:=0; EMPTY:=1;
4600.      AT(/3/):=ATMOD+DELAY;
4700.      'GOTO' AMI;
4800.
4900.AMT3: 'COMMENT' END OF PROCESSING;
5000.      IDLE:=1;
5100.      AO(/INNR/):=0;
5200.      'IF' FULL=1 'THEN' 'GOTO' AMT1;
5300.      'GOTO' AMV;
5400.
5500.'END' BOX 1;
5600.
5700.$MEND;
```

a

```
100.$MACRO BQ2;
200.$DC 23;
300.$DC 29;
400.$DC 1;
500.'BEGIN'
600.'COMMENT' TERMINAL TO BE CONNECTED TO PROCESSOR BQ1,
700.          AT(/1/) : PRE-THINKING AT THE TERMINAL,
800.          AT(/2/) : TYPING,
900.          AI(/1/) : 1 = RESPONSE FROM PROCESSOR;
1000.'OWN' 'REAL' OXX,OYY,RO,RI;
1100.'OWN' 'INTEGER' IAI,WAITDI;
1200.'OWN' 'BOOLEAN' POISSWI;
1300.'BOOLEAN' FEHLER;
1400.'INTEGER' T;
1500.
1600.'PROCEDURE' THINK;
1700.'BEGIN'
1800.   RANERL (1/50,1,OXX,T,FEHLER);
1900.   AT(/1/):=ATMOD+T;
2000.'END' THINK;
2100.
2200.'SWITCH' AMSTART:=AMNR,AMSE,AMOP,AMT1,AMT2,AMI1;
2300.'GOTO' AMSTART(/AJX/);
2400.
2500.AMNR: ATZ:=2; AIZ:=1; AOZ:=1; 'GOTO' AMV;
2600.
2700.AMSE: PAMSE (ABN);
2800.      OXX:=720+ABN; POISSWI:='TRUE';
2900.      IAI:=0;
3000.      'COMMENT' START OF PRE-THINKING;
3100.      THINK;
3200.      'GOTO' AMV;
3300.
3400.AMT1: 'COMMENT' PRE-THINKING FINISHED;
3500.      WAITDI:=ATMOD;
3600.      RAPOIS (30,POISSWI,OYY,OXX,RO,FEHLER);
3700.      RI:=1;
3800.      'COMMENT' START OF TYPING;
3900.      'IF' IAI=0 'THEN' AT(/2/):=ATMOD+20;
4000.      'GOTO' AMV;
4100.
4200.AMI1: 'COMMENT' RESPONSE FROM PROCESSOR;
4300.      'IF' AI(/1/)=1
4400.      'THEN' 'BEGIN'
4500.      AO(/1/):=0; IAI:=1;
4600.      'END' THEN
4700.      'ELSE' 'BEGIN'
4800.      IAI:=0;
4900.      'IF' RI=1
5000.      'COMMENT' START OF TYPING;
5100.      'THEN' AT(/2/):=ATMOD+20;
5200.      'END' ELSE;
5300.      'GOTO' AMV;
5400.
5500.AMT2: 'COMMENT' TYPING FINISHED, SENDING OF SIGNAL TO PROCESSOR;
5600.      WAIT (WAITDI,ABN);
5700.      AO(/1/):=RO; STATISTICS;
5800.      'COMMENT' START OF PRE-THINKING;
5900.      THINK;
6000.      RI:=0;
6100.      IAI:=2;
6200.      'GOTO' AMV;
6300.
6400.'END' BOX 2;
6500.$MEND;
```

b

◁ **Fig. 16.** BOMOL program obtained from the flow-chart description of figs 14 and 15. *a*) The central processor ('MACRO BQ1'). *b*) A terminal ('MACRO BQ2').

```
100.$MACRO NQ1;
200.
300.CR 1.1.1=2.1.1;
400.CR 1.2.1=2.1.2;
500.CR 1.3.1=2.1.3;
600.CR 1.4.1=2.1.4;
700.CR 2.1.1=1.1.1;
800.CR 2.1.2=1.2.1;
900.CR 2.1.3=1.3.1;
1000.CR 2.1.4=1.4.1;
1100.
1200.$MEND;
```

```
100.$MACRO NQ2;
200.
300.CR 1.1(1--4) = 2(1--4)1;
400.CR 2(1--4)1  = 1.1(1--4);
500.
600.$MEND;
```

Fig. 17. Description in COMOL of the connection network of the system shown in fig. 13. The measuring instrument has been omitted. *Top:* Full description. *Bottom:* Equivalent abbreviated notation.

transition time is 2 units. Third, the actual processing of the job by the central processor; the transition time here is a variable quantity, and carried as an attribute with the token. Fourth, the branching of signals from the processor to the terminal; no transition time is involved here, and the destination is carried with the token as an attribute.

There are two different kinds of transition for the terminal: the first represents the generation of a new job; the duration of this step is taken from a suitable random process ('ranerl'). The second, numbered *3* in the diagram, represents the typing actions for introducing the job into the central processor; the transition time for this is fixed and equal to 20 units. The transition numbered *2* in the diagram has no significance but is required because location *2* cannot be input and output of the same transition at the same time. *Fig. 19* shows the SAMEN description in program form.

A comparison of the E-net description and the SAMO description, where we find that the thinking period does not start before the typing has finished, might suggest that things are different here. However, it should be borne in mind that when a transition, e.g. transition *3*, fires, the token in location *3* moves on to *5* after the transition time has elapsed and this transition time starts counting at the moment when input tokens are present and output locations have become empty. So it is also true that in this E-net description the thinking period cannot start before the typing has finished.

As mentioned before, one of the major frustrations in using a computer is having to wait to type in a new job because the central processor has not yet finished the previous one. The main result of the present example is therefore a list of minimum, maximum and average waiting times and the number of jobs done in

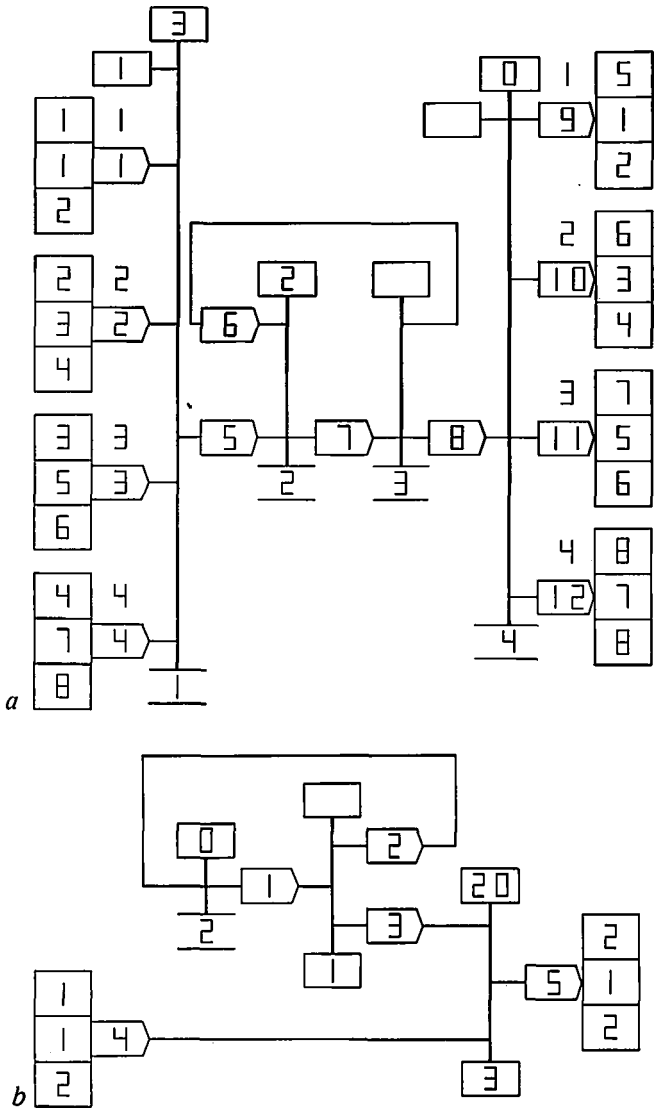


Fig. 18. *a*) Evaluation net for the central processor (fig. 13) as drawn with the aid of the GIPSIE program. The transitions shown are: *1* arrival of job from a terminal and reading information from disc. *2* loading job into memory. *3* execution of job. *4* branching to terminal.
b) Evaluation net of a terminal. The transitions introduced here are: *1* terminal generates new job. *2* no action (see text). *3* typing action for sending a job to the central processor.

40 time units for the first three terminals. As can clearly be seen, there is a jump between the waiting times of terminal *4* and the other terminals, and this demonstrates that the processor is overloaded by four terminals under the given conditions. As a consequence of these simulation results it would be necessary to make the central processor faster or to reduce the number of terminals connected to it.

```
100.$MACRO BQ1001;
200.'BEGIN'
300.'INTEGER''PROCEDURE' R4;
400.  R4:=B8(1);
500.'INTEGER''PROCEDURE' T3;
600.  T3:=B7(2);
700.AMSE: BTOAT(6,2);
800.IN1:  BIN   (1,1,2);
900.IN2:  BIN   (2,3,4);
1000.IN3: BIN   (3,5,6);
1100.IN4: BIN   (4,7,8);
1200.OUT5: BOUT  (9,1,2);
1300.OUT6: BOUT  (10,3,4);
1400.OUT7: BOUT  (11,5,6);
1500.OUT8: BOUT  (12,7,8);
1600.DA1:  BYNTRA(1,2,3,4,5,60,Q0,1);
1700.DA2:  BJTRA (5,6,7,80,Q0);
1800.DA3:  BFTRA (7,8,6,T3,Q0);
1900.DA4:  BXNTRA(8,9,10,11,12,4,Q0,R4);
2000.'END';
2100.$MEND;
```

a

```
100.$MACRO BQ1002;
200.$DC 23;
300.$DC 29;
400.$DC 1001;
500.'BEGIN'
600.'OWN''REAL'XX,XXX; 'OWN''BOOLEAN'POISSWI;
700.'BOOLEAN' FEHLER;
800.
900.'INTEGER''PROCEDURE' T1;
1000.'BEGIN' 'INTEGER' T;
1100.  RANERL(1/300,1,XX,T,FEHLER);
1200.  T1:=T;
1300.'END';
1400.
1500.'PROCEDURE'Q1;
1600.'BEGIN'
1700.  B3(1):=B1(1);
1800.  RAPOIS(100,POISSWI,XXX,XX,B3(2),FEHLER);
1900.  SET TSTART;
2000.'END';
2100.
2200.'PROCEDURE' Q3;
2300.'BEGIN' COMPUTE STATISTICS; 'END';
2400.
2500.AMSE: XX:=7200+ABN;
2600.POISSWI:='TRUE';
2700.INITIALIZATION;
2800.BTOAT(3,2);
2900.
3000.IN1:  BIN   (4,1,2);
3100.OUT2: BOUT  (5,1,2);
3200.DA1:  BFTRA (1,2,3,T1,Q1);
3300.DA2:  BTTRA (2,1,0,Q0);
3400.DA3:  BJTRA (3,4,5,150,Q3);
3500.'END';
3600.$MEND;
```

b

```
100.$MACRO NQ1001;
200.CB 1002(1--4)2=1001.1(1--4);
300.CB 1001.1(1--4)=1002(1--4)2;
400.CB 1001.1(5--8)=1002(1--4)1;
500.CB 1002(1--4)1=1001.1(5--8);
600.CR 1002(1--4)1=1001.1(1-3-5-7);
700.CR 1002(1--4)2=1001.1(2-4-6-8);
800.CR 1001.1(1-3-5-7)=1002(1--4)1;
900.CR 1001.1(2-4-6-8)=1002(1--4)2;
1000.$MEND;
```

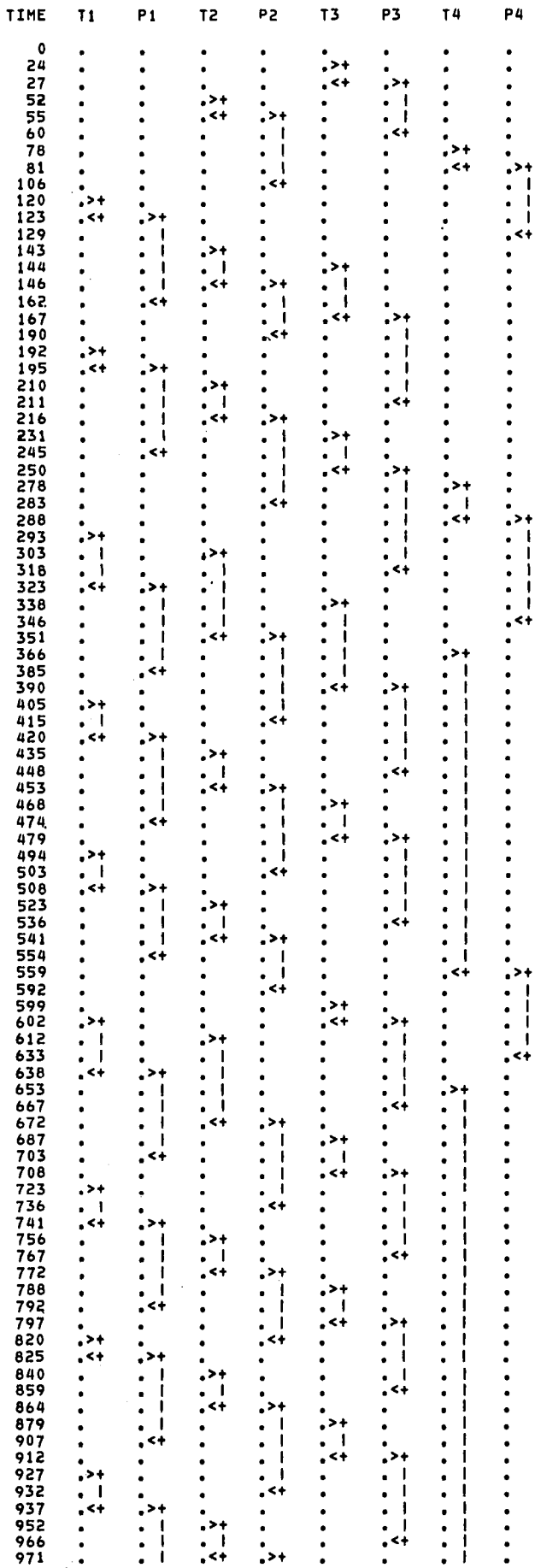
c

Fig. 19. SAMEN description of central processor (*a*), terminal (*b*) and connection net (*c*). (See also fig. 18.)

subsequent periods of about 1000 time units, together with the activity diagram, for each terminal; see *fig. 20*. These results show that the maximum waiting time increases slightly with decreasing priority of the terminal (terminal *1* has the highest priority), while the average waiting time is nearly constant at about

For checking a program for the first time, complete step-by-step print-outs of the simulation run can be obtained in SAMO as well as in SAMEN programs, and we have used these in constructing the diagram.

In their detailing, both SAMO and SAMEN programs produce exactly the same results. The SAMEN descrip-




```
100.
200.WAITING TIME OF THE TERMINALS:
300.TERMINAL  MINIMUM  MAXIMUM  AVERAGE  JOBS  TIME=   1008
400.      1        0       95       26     9
500.      2        0      110       47    10
600.      3        0       58       18    10
700.      4        0      242       65     4
800.
900.
1000.WAITING TIME OF THE TERMINALS:
1100.TERMINAL  MINIMUM  MAXIMUM  AVERAGE  JOBS  TIME=   2020
1200.     1        0      105       38    19
1300.     2        0      110       36    20
1400.     3        0       75       26    18
1500.     4        0      463      169     8
1600.
1700.
1800.WAITING TIME OF THE TERMINALS:
1900.TERMINAL  MINIMUM  MAXIMUM  AVERAGE  JOBS  TIME=   3020
2000.     1        0      105       36    30
2100.     2        0      110       33    28
2200.     3        0      110       36    27
2300.     4        0      463      154    12
2400.
2500.
2600.WAITING TIME OF THE TERMINALS:
2700.TERMINAL  MINIMUM  MAXIMUM  AVERAGE  JOBS  TIME=   4002
2800.     1        0      105       37    38
2900.     2        0      110       41    38
3000.     3        0      128       40    36
3100.     4        0      529      185    15
3200.
3300.
3400.WAITING TIME OF THE TERMINALS:
3500.TERMINAL  MINIMUM  MAXIMUM  AVERAGE  JOBS  TIME=   5015
3600.     1        0      105       39    48
3700.     2        0      110       44    46
3800.     3        0      128       41    43
3900.     4        0      570      171    20
4000.
4100.
4200.WAITING TIME OF THE TERMINALS:
4300.TERMINAL  MINIMUM  MAXIMUM  AVERAGE  JOBS  TIME=   6028
4400.     1        0      105       39    57
4500.     2        0      110       41    55
4600.     3        0      128       39    50
4700.     4        0      570      137    27
4800.
4900.
5000.WAITING TIME OF THE TERMINALS:
5100.TERMINAL  MINIMUM  MAXIMUM  AVERAGE  JOBS  TIME=   7003
5200.     1        0      105       38    65
5300.     2        0      110       40    62
5400.     3        0      128       39    57
5500.     4        0      570      121    34
```

Fig. 20. Results of the simulation of the system of fig. 13. *On the left:* Activity diagram produced by the simulation program for the same situation as shown in diagram I. Columns *T1* to *T4* show the times at which information is sent to the central processor (> +) and the times at which the processor has accepted the information and finished reading the additional information from disc into the buffer (< +). Columns *P1* to *P4* show the times at which the central processor starts loading from the buffer into the main memory for subsequent processing (> +) and the times at which the processor has finished the job (< +). *Above:* quantitative data for each terminal relating to minimum, maximum and average waiting times and the number of jobs executed during a part of the simulation.

tion, since it is a somewhat later development, may be a little shorter and easier to use in most cases; however, for the simulation of electronic circuits at the most detailed level SAMO may offer greater scope.

The programs described in this article are available in various parts of the Philips group for P1400 computers in combination with a Multiprogramming Dialogue operating System (MDS). Implementation on other computers is in progress. The method has been used to verify logic designs at gate level and to verify the design of pulse-code modulation system components at bit level. It has been used to investigate certain computer hardware concepts at register-transfer level. SAMO has been used by students to simulate rail traffic. However, one of the main fields of application was in the project 'Optimization of computer net-

works with terminals' [8–12], a joint venture with Philips GmbH Data Systems, Werk für Daten- und Informationstechnik, Siegen; an example is presented in fig. 13. Current usage includes the analysis of system concepts for computer-controlled telecommunication switching systems.

[8] J. Schwandt, An approach to use evaluation nets for the performance evaluation of transaction-oriented business computer systems, in: K. M. Chandy and M. Reiser (eds), Computer performance, pp. 429-441, North-Holland, Amsterdam 1977.

[9] H. Seidel and P. v. Studnitz, A methodology for deterministic model building and simulation of computer systems, Simulation '77, Proc. Conf. Montreux 1977, pp. 55-60.

[10] H. Seidel and P. v. Studnitz, Modellbildung und Simulation kommerzieller Anwendungen, Angew. Informatik **21**, 108-117, 1979.

[11] R. Winter, An evaluation net model for the performance evaluation of a computer network, in: H. Beilner and E. Gelenbe (eds), Measuring, modelling, and evaluating computer systems, pp. 95-113, North-Holland, Amsterdam 1977.

[12] J. Schwandt, Interaktiver rechnerunterstützter Systementwurf, VDI-Z **121**, 559-564, 1979.

Summary. Methods of description for time-discrete systems have been developed at Philips Forschungslaboratorium Hamburg and used for the simulation of such systems, resulting in the SAMEN/SAMO package. To keep the system description clear and distinct and yet as versatile as possible, it is divided into subsystems with their interconnections. The subsystems can be treated as individual 'black boxes' or further subdivided as required. At the more general SAMO level of the SAMEN/SAMO program package the description of the subsystems follows a convention BOMOL (Box-Model Description Language), whereas the connections are described in COMOL notation (Connection-Model Description Language). To provide additional support in model description evaluation nets ('E-nets') were chosen as a formal description method, and a 'SAMEN level' was added to the simulation-program package. The SAMEN description of the black boxes, based on evaluation nets, can replace the BOMOL part of the simulation program. E-nets not only have the advantage that they provide the user with a clear insight, in the form of a drawing, into the structure of his problem, they can also be clearly described in a formal manner. This is essential when producing the simulation program. For communication with the computer there is a set of programs called IMOBS (Interactive Model Building and Simulation), consisting of a program package GIPSIE (Graphical Interactive Program for Simulation with Evaluation Nets) that can be used to build up an evaluation net on a graphic-display terminal, and a program package DIMPLE (Dialogue Implementation Package) that can be used in a dialogue with the computer to enter values of the model parameters and prepare the simulation.

Scientific publications

These publications are contributed by staff of laboratories and plants which form part of or cooperate with enterprises of the Philips group of companies, particularly by staff of the following research laboratories:

Philips Research Laboratories, Eindhoven, The Netherlands	*E*
Philips Research Laboratories, Redhill, Surrey, England	*R*
Laboratoires d'Electronique et de Physique Appliquée, 3 avenue Descartes, 94450 Limeil-Brévannes, France	*L*
Philips GmbH Forschungslaboratorium Aachen, Weißhausstraße, 51 Aachen, Germany	*A*
Philips GmbH Forschungslaboratorium Hamburg, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany	*H*
Philips Research Laboratory Brussels, 2 avenue Van Becelaere, 1170 Brussels (Boitsfort), Belgium	*B*
Philips Laboratories, N.A.P.C., 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A.	*N*

H. Baudry & F. Franconville (CII-Honeywell Bull, St-Ouen, France): Encres sérigraphiables pour haute définition. Rhéologie et impression.
Acta Electronica **21**, 283-295, 1978 (No. 4). *L*

J. van den Boomgaard: Eutectic-like and peritectic-like reactions in ternary systems at a constant pressure, Part III. Ternary systems containing more than three solid phases.
Philips J. Res. **34**, 211-237, 1979 (No. 5/6). *E*

J. van den Boomgaard & R. A. J. Born: A sintered magnetoelectric composite material $BaTiO_3$-Ni(Co, Mn)Fe_2O_4.
J. Mat. Sci. **13**, 1538-1548, 1978 (No. 7). *E*

C. H. J. van den Brekel: Mass transport and morphology in chemical vapour deposition processes.
Thesis, Nijmegen 1978. *E*

H. H. Brongersma, N. Hazewindus, J. M. van Nieuwland, A. M. M. Otten & A. J. Smets: 'NODUS' — a sensitive new instrument for analysing the composition of surfaces.
Rev. sci. Instr. **49**, 707-714, 1978 (No. 6). *E*

A. Browne: A reading aid for the partially sighted.
Electronics & Power **24**, 592-595, 1978 (Aug.). *R*

P. J. Cameron (Merton College, Oxford), **J.-M. Goethals & J. J. Seidel** (Eindhoven University of Technology): The Krein condition, spherical designs, Norton algebras and permutation groups.
Proc. Kon. Ned. Akad. Wetensch. A **81**, 196-206, 1978 (No. 2). *B*

W. A. Cense & W. Albers: On the in-situ growth of aligned composites by means of disproportionation of amorphous metal systems.
J. non-cryst. Solids **28**, 391-401, 1978 (No. 3). *E*

T. A. C. M. Claasen & W. F. G. Mecklenbräuker: A generalized scheme for an all-digital time-division multiplex to frequency-division multiplex translator.
IEEE Trans. **CAS-25**, 252-259, 1978 (No. 5). *E*

H. J. van Daal, K. H. J. Buschow & P. B. van Aken: On the thermoelectricity in uranium intermetallics.
Thermoelectricity in metallic conductors, ed. F. J. Blatt & P. A. Schroeder, pp. 107-115; Plenum Press, New York 1978. *E*

N. H. Dekkers & H. de Lang: A calculation of bright field single-atom images in STEM with half plane detectors.
Optik **51**, 83-92, 1978 (No. 1). *E*

Ph. Delsarte & Ph. Piret: Automorphism groups of convolutional codes.
SIAM J. appl. Math. **34**, 616-629, 1978 (No. 4). *B*

O. P. van Driel: On various causes of improper solutions in maximum likelihood factor analysis.
Psychometrika **43**, 225-243, 1978 (No. 2). *E*

A. van der Drift (Philips Centre for Technology, Eindhoven), **G. W. Griffiths** (Newmarket Transistors Ltd, Newmarket, Suffolk, U.K.) **& R. Petit** (Portenseigne, Louviers, France): Mounting technologies for hybrid integrated circuits and their mechanization.
Acta Electronica **21**, 297-318, 1978 (No. 4). *(Partly in English, partly in French.)*

J. M. Durand: Influence of the growth parameters in GaAs vapour phase epitaxy.
Philips J. Res. **34**, 177-210, 1979 (No. 5/6). *L*

A. van Eenbergen & E. Bruninx: Losses of elements during sample decomposition in an acid-digestion bomb.
Anal. chim. Acta **98**, 405-406, 1978 (No. 2). *E*

U. Enz: A particle model based on stringlike solitons.
J. math. Phys. **19**, 1304-1306, 1978 (No. 6). *E*

J. van Esdonk: Verbindingstechnieken voor ultrahoog-vacuümsystemen.
Lastechniek **44**, 53-60, 1978 (No. 4). *E*

E. Fischmann: Alterations in the manuscript of Röntgen's publication 'Ueber eine neue Art von Strahlen' (1895).
Brit. J. Radiol. **51**, 475-476, 1978 (No. 606). *E*

W. P. J. Fontein, J. M. M. Verbakel & J. H. N. van Vucht: A trial-structure generating program for the determination of crystal structures.
Philips J. Res. **34**, 238-251, 1979 (No. 5/6). *E*

C. T. Foxon & B. A. Joyce: Surface processes controlling the growth of $Ga_xIn_{1-x}As$ and $Ga_xIn_{1-x}P$ alloy films by MBE.
J. Crystal Growth **44**, 75-83, 1978 (No. 1). *R*

W. Funk: Thick film technology.
Acta Electronica **21**, 251-255, 1978 (No. 4). *H*

J. Georges & F. Heymans: On compound statements in a conversational language.
BIT **18**, 230-233, 1978 (No. 2). *B*

J. T. H. Goossen & J. G. Kloosterboer: Determination of phosphates in natural and waste waters after photochemical decomposition and acid hydrolysis of organic phosphorus compounds.
Anal. Chem. **50**, 707-711, 1978 (No. 6). *E*

H.-J. Hagemann: Loss mechanisms and domain stabilisation in doped $BaTiO_3$.
J. Physics C **11**, 3333-3344, 1978 (No. 15). *A*

J. C. M. Henning & J. H. den Boef: Magnetostriction measurement by means of strain modulated ferromagnetic resonance (SMFMR).
Appl. Phys. **16**, 353-357, 1978 (No. 4). *E*

E. P. Honig & B. R. de Koning: Transient and alternating electric currents in thin organic films.
J. Physics C **11**, 3259-3271, 1978 (No. 15). *E*

H. Ihrig: The positive temperature coefficient resistivity of $BaTiO_3$ ceramics as a function of the amount of titanium-rich second phase.
Phys. Stat. sol. (a) **47**, 437-444, 1978 (No. 2). *A*

G. Jacob, M. Boulou, M. Furtado & D. Bois: Gallium nitride emitting devices. Preparation and properties.
J. electronic Mat. **7**, 499-514, 1978 (No. 4). *L*

F. de Jager & C. B. Dekker: Tamed frequency modulation, a novel method to achieve spectrum economy in digital transmission.
IEEE Trans. **COM-26**, 534-542, 1978 (No. 5). *E*

D. J. Kroon: Instrumental methods for automatic air monitoring stations.
Air pollution control, Part III, ed. W. Strauss, pp. 317-386; Wiley, New York 1978. *E*

A. R. Miedema: Surface segregation in alloys of transition metals.
Z. Metallk. **69**, 455-461, 1978 (No. 7). *E*

M. Monneraye: Les encres sérigraphiables en microélectronique hybride: les matériaux et leur comportement.
Acta Electronica **21**, 263-281, 1978 (No. 4). *L*

M. Monnier: Une nouvelle technique de montage et d'interconnexion de composants: circuiterie double face sur film polyimide.
Acta Electronica **21**, 333-341, 1978 (No. 4). *L*

J. H. C. van Mourik (Philips Centre for Technology, Eindhoven): Screen printable inks in hybrid microelectronics, availability and physical properties.
Acta Electronica **21**, 257-262, 1978 (No. 4).

B. J. Mulder: Protective glassy layers passivating copper at 500 °C.
J. Electrochem. Soc. **125**, 823-824, 1978 (No. 5). *E*

R. Orlowski & E. Krätzig: Holographic method for the determination of photo-induced electron and hole transport in electro-optic crystals.
Solid State Comm. **27**, 1351-1354, 1978 (No. 12). *H*

F. Perrault (T.R.T., Le Plessis-Robinson, France), **B. Kramer, R. Dessert & G. Le Floch:** Applications des technologies couches minces et couches épaisses dans le domaine des hautes fréquences.
Acta Electronica **21**, 343-354, 1978 (No. 4). *L*

E. Roza: A practical design approach to decision feedback receivers with conventional filters.
IEEE Trans. **COM-26**, 679-689, 1978 (No. 5). *E*

H. Schauer, E. Arnold & P. C. Mürau: Interface states and memory decay in MNOS capacitors.
IEEE Trans. **ED-25**, 1037-1041, 1978 (No. 8). *N*

C. Schiller, A. Gheorghiu* & M. L. Theye* (* Université Pierre et Marie Curie, Paris): Etude de la composition de couches minces de GaAs amorphe par analyse non dispersive de rayons X.
J. Microsc. Spectrosc. électron. **3**, 255-263, 1978 (No. 3). *L*

H. Sewell: Control of pattern dimensions in electron lithography.
J. Vac. Sci. Technol. **15**, 927-930, 1978 (No. 3). *R*

B. Singer & A. L. Dalisa: An x-y addressable electrophoretic display.
Proc. S.I.D. **18**, 255-266, 1977 (No. 3/4). *N*

A. L. N. Stevels: Effect of non-stoichiometry on the luminescence of Eu^{2+}-doped aluminates with the β-alumina-type crystal structure.
J. Luminescence **17**, 121-133, 1978 (No. 1). *E*

N. C. de Troye: Integration durch Integration.
E und M **95**, 287-292, 1978 (No. 6/7). *E*

V. Tulaja (Fernuniversität, Hagen), **B. Schiek & J. Köhler:** An interdigitated 3-dB coupler with three strips.
IEEE Trans. **MTT-26**, 643-645, 1978 (No. 9). *H*

R. Veilex: Introduction *(to issue on Hybrid microelectronics)*.
Acta Electronica **21**, 247-249, 1978 (No. 4). *(In English and in French.)*

H. W. Werner: Introduction to secondary ion mass spectrometry (SIMS).
Electron and ion spectroscopy of solids, ed. L. Fiermans, J. Vennik & W. Dekeyser, pp. 324-441; Plenum Press, New York 1978. *E*

G. F. Weston: Pumps for ultra-high vacuum.
Vacuum **28**, 209-233, 1978 (No. 5). *R*

Contents of Philips Telecommunication Review **37**, No. 3, 1979:

G. Toes: TSAPT for automatic public telegraphy (pp. 97-110).
D. van der Lugt & J. van Zweeden: Private automatic text and telex exchange DSX-40 (pp. 111-124).
W. F. Njio & B. van Raay: The 8TR 400 family of supergroup translating equipment (pp. 125-135).
W. F. Njio & A. Klaassen: Compact 8TR 417 master frequency generating equipment (pp. 136-143).
W. G. Bax & J. Wagenmakers: 140 Mb/s coaxial transmission system 8TR 609 (pp. 144-160).
H. W. de Haan & P. Zwaal: 34 Mb/s line system 8TR 607 (pp. 161-169).
G. David & B. Pando: An on-board digital transmission system (pp. 170-176).
T. G. Giles: A universal frequency synthesizer IC (pp. 177-181).
F. van Enk: Microprocessor control of mobile radio systems (pp. 182-187).
M. Pirson: A new vehicular and manpack HF/SSB radio (pp. 188-195).

Contents of Philips Telecommunication Review **37**, No. 4, 1979:

K. Mouthaan: Telecommunication via glass-fibre cables (pp. 201-214).
J. W. Versluis & J. G. J. Peelen: Optical communication fibres. Manufacture and properties (pp. 215-230).
F. Krahn, G. Olejak & W. Weidhaas: The manufacture of optical cables (pp. 231-240).
F. Krahn, W. Meininghaus & D. Rittich: Measuring and test equipment for optical cable (pp. 241-249).
O. R. Bresser: Metal-free optical fibre cables (pp. 251-256).
F. Krahn & H. Thielman: 34 Mb/s optical cable system test in Berlin (pp. 257-269).

Contents of Electronic Components and Applications **1**, No. 4, 1979:

F. A. M. van de Kerkhof: Microcomputer-controlled tuning and control systems for TV (pp. 199-214).
A. Lieders: Single-phase rectifier circuits with CR filters, Part 2 — Design procedure (pp. 216-230).
J. W. Beunders: Digital control of radio and audio equipment, Part 2 — Voltage converter and memory for preset radio tuning (pp. 231-238).
U. Schillhof: Digital control of radio and audio equipment, Part 3 — Introduction to microcomputer-controlled radio tuning system (pp. 239-248).
E. D. van Veldhuizen: The 8X300 development and monitoring system (DMS) (pp. 249-257).

Contents of Electronic Components and Applications **2**, No. 1, 1979:

U. Schillhof: Digital control of radio and audio equipment, Part 4 — Displays and control of analogue functions with RTS (pp. 2-15).
E. A. Kilian: Digital control of radio and audio equipment, Part 5 — DC-controlled audio switches and filters (pp. 16-26).
A new approach to a PCM codec for telephony (pp. 28-30).
H. Houkes: TDA1060 — a comprehensive integrated control circuit for SMPS (pp. 31-48).
W. Ebbinge: Mains pollution caused by domestic appliances, Part 1 — Definitions and standards (pp. 49-52).
P. Blunt: Reliable thyristors and triacs in TO-220 plastic packages (pp. 53-58).
Segment field magnets for large electric motors (pp. 59-61).

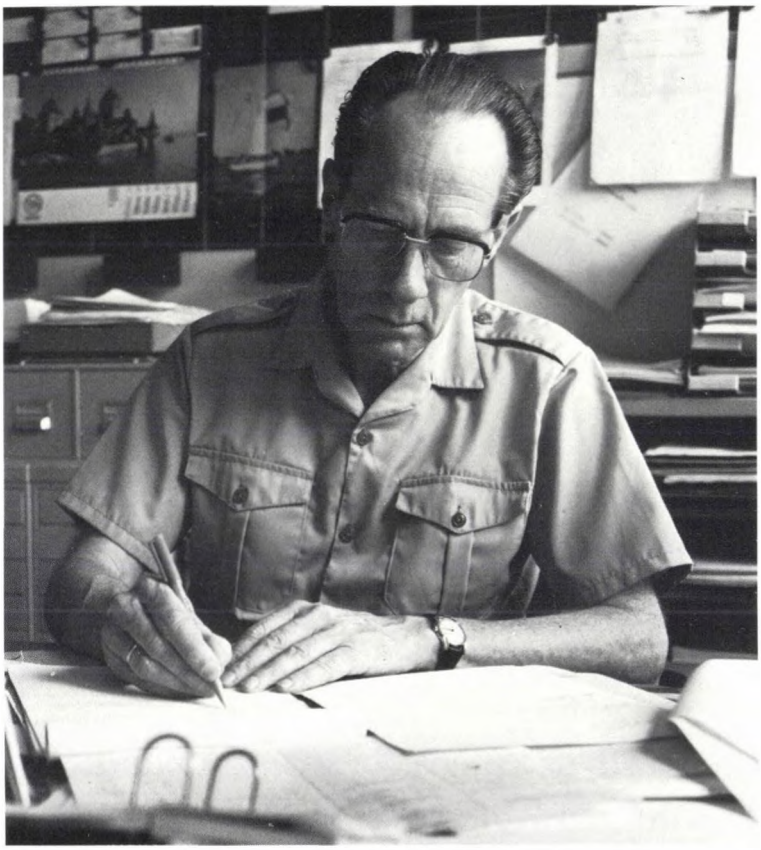
Contents of Valvo Berichte **20**, No. 3, 1979:

Das pyroelektrische Vidikon (pp. 93-112).
P. Dolizy, O. De Luca & M. Deloron: Fotoemission von multialkalischen Antimonlegierungen vom Typ (Sb,Na,K,Cs) (pp. 113-146).

Volume 39, 1980, No. 5 pages 125-152 Published 3rd December 1980

PHILIPS TECHNICAL REVIEW

VOLUME 39, 1980, No. 6/7



On 30th September 1980 Drs J. W. Miltenburg retired from Philips Technical Review. He had been with the Review since 1st January 1957, and succeeded Ir Gradstein as Editor-in-chief in 1970.

A notable character thus leaves the stage. A character who thrived on hard work and was fascinated by all the aspects of producing a journal: not just the contents, but the illustrations, typography and design as well. On these matters, and especially on editing, he had his own views, often strong ones. This meant that those who helped in producing the Review did not always have an easy time. But neither did he spare himself in his unrelenting struggle to maintain the quality of the journal. The proper use of the Dutch language was particularly close to his heart.

During Miltenburg's time it became customary to conclude each volume with a double issue entirely devoted to some special Philips area of expertise. The five most recent of these double issues treated inorganic chemical analysis, light, modulation, LSI and picture processing. These special issues required a great deal of organization and coordination, and Miltenburg always shouldered these responsibilities with energy and enthusiasm.

So now we offer Drs Miltenburg — in 'his' journal — our best wishes for the future. We hope that his successor and her colleagues will cherish the Philips Technical Review with the same care and affection.

E. F. de Haan

The deflection coils of the 30AX colour-picture system

W. A. L. Heijnemans, J. A. M. Nieuwendijk and N. G. Vink

The 30AX system, which Philips introduced in 1978, is an important landmark in the development of colour-picture systems. With previous systems the assembly technician had to work through a large number of complicated setting-up procedures whenever he fitted a television picture tube with a set of coils for deflecting the electron beams. These procedures were necessary to ensure that the beams for the three colours would converge at the screen for every deflection. They are no longer necessary with the 30AX system: for a given screen format any deflection unit can be combined with any tube to form a single 'dynamically convergent' unit. A colour-television receiver can thus be assembled from its components almost as easily as a monochrome receiver.

In their work on the design of deflection coils in the last few years the authors have expanded the magnetic deflection fields into 'multipoles'. This approach has improved the understanding of the relations between coil and field and between field and deflection to such an extent that designing deflection units is now more like playing a difficult but fascinating game of chess than carrying out the obscure computing procedure once necessary.

Introduction

All colour-television receivers sold by Philips now have a shadow-mask tube. The red, green and blue pictures on the screen of such a tube are traced out by three separate electron beams *R*, *G* and *B* (*fig. 1a*); the beam *R* only strikes the 'red' phosphor on the screen, *G* only the 'green' one and *B* only the 'blue' one (fig. 1*b*).

The deflection unit (*A* in fig. 1*a*) consists primarily of a line coil for the horizontal deflection and a frame coil for the vertical deflection. For the system to be 'dynamically convergent' — i.e. for the three beams to converge at the screen for every deflection — the deflection unit has to satisfy very exacting electron-optical requirements. To meet the standards we have

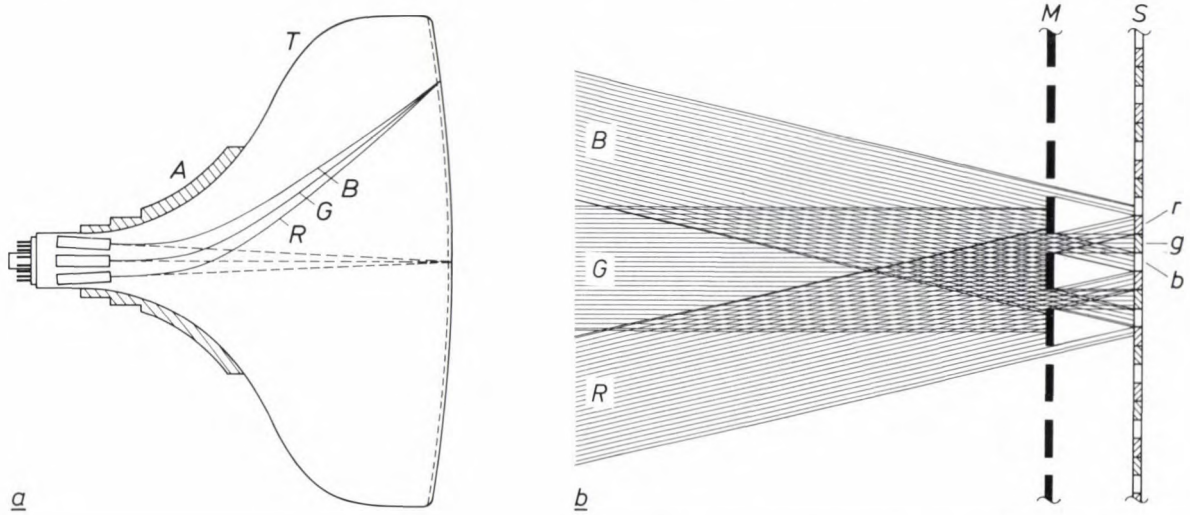


Fig. 1. Shadow-mask tube. *a*) The electron beams *R*, *G* and *B* converge at the screen of tube *T*; they are controlled by a line coil and a frame coil, which together form the deflection unit *A*. *b*) Because of the shadowing effect of the mask *M*, beam *R* only strikes the 'red' phosphor *r* on the screen *S*, *G* only the 'green' phosphor *g* and *B* only the 'blue' phosphor *b*.

Ir W. A. L. Heijnemans is with Philips Research Laboratories, Ir J. A. M. Nieuwendijk is with Philips Audio Division and Ing. N. G. Vink is with Philips Video Division, all in Eindhoven.

adopted, the landing points must never lie more than 0.5 mm apart.

In this article we shall discuss the deflection unit of the 30AX colour-picture system. First of all we shall briefly describe the 30AX system itself and the systems that preceded it: the delta systems and the 20AX system. *Fig. 2* shows the deflection units for the 20AX and the 30AX systems in position on the tube.

In a delta system the three electron guns are arranged in a triangle (or 'delta'). A disadvantage of this apparently obvious arrangement is that it is impossible to design a 'four-wire unit' — a deflection unit with only four connecting wires, two for the line coil and two for the frame coil — that will make the beams converge on the screen for every deflection. The dynamic convergence in these systems is produced by an additional set of coils supplied by currents related to the deflection currents in a complicated way. Setting up these currents is a particularly tedious procedure: no fewer than 14 electronic settings are necessary, and they are all interdependent. Because of these complications, the use of delta systems is decreasing.

Convergence *can* however be obtained with a four-wire unit if the electron guns are side by side in the same plane (*fig. 3*). A closer analysis of such an 'in-line' system shows that it offers simpler solutions in all respects [1]. In 1974 Philips introduced the 20AX system, which is based on this principle. It contains only one additional coil for correcting residual errors.

Even with the 20AX system, however, the assembly technician, every time he combines a deflection unit with a tube, has to carry out a large number of complicated operations to obtain a picture with true colours and no distortion. The necessity for this arises mainly from the spread in the dimensions and other characteristics of the tubes and coils when they come from the factory.

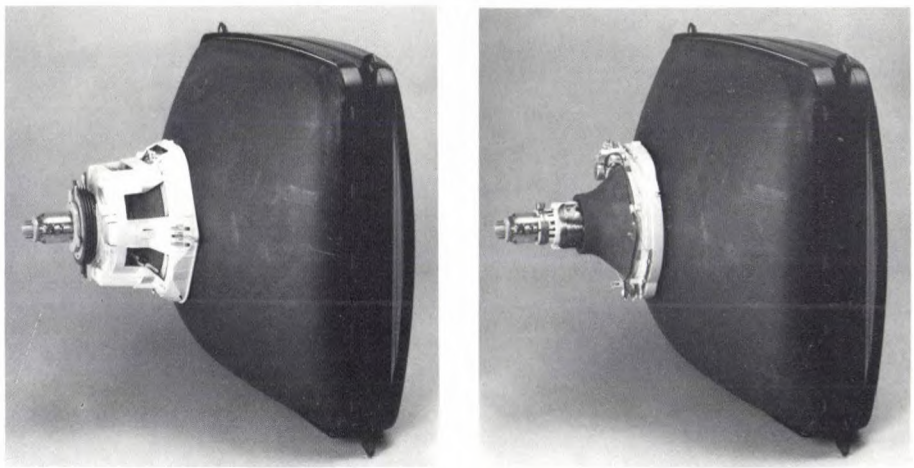


Fig. 2. The deflection units for the 20AX system (*left*) and for the 30AX system (*right*). The photograph on the left shows in addition to the deflection unit — the coil assembly in a plastic encapsulation — a dark ring with projections; this is used to produce 'static convergence' in the tube, i.e. convergence in the absence of deflection fields. In the 30AX system the tube is already 'statically convergent' when it is delivered from the tube factory, i.e. before the deflection unit is mounted.

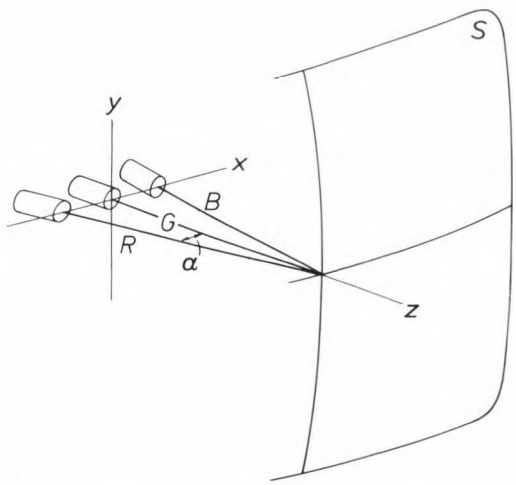


Fig. 3. 'In-line' system. The guns for the electron beams are side by side in the same plane. In the undeflected state, beams *R*, *G* and *B* lie in a single plane, which is the horizontal plane in all known systems (in this article it is the *x*,*z*-plane), and converge at the centre of the screen *S*.

[1] J. Haantjes and G. J. Lubben, Philips Res. Repts. **12**, 46, 1957, and **14**, 65, 1959.

The 30AX system

The 30AX system is based on the idea of first making a design that intrinsically has no deflection errors, and then making the tubes and coils to this design in the factory so accurately that corrections during assembly are unnecessary. With this system all the technician has to do for a given type of screen is to slide any deflection unit over the neck of any tube and fix it in position; he then has a completely 'self-convergent' assembly, i.e. one that is dynamically convergent with no need for any setting up at all. The deflection unit and the tube fit together by means of three reference bosses that guarantee the correct relative positions (*fig. 4*).

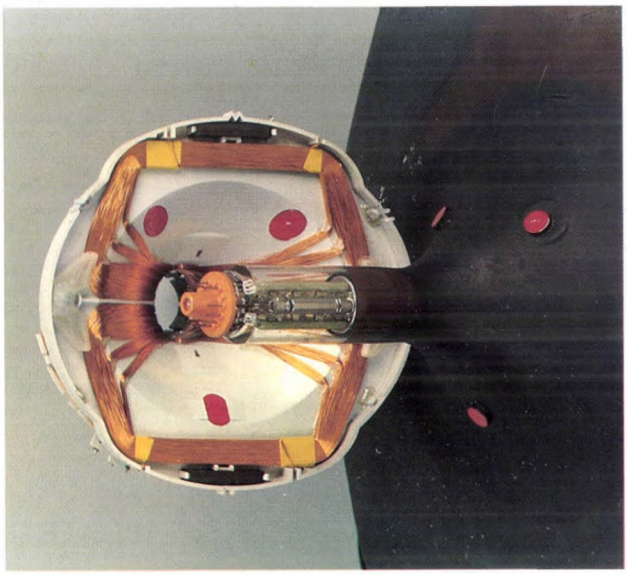


Fig. 4. The 30AX deflection unit seen from inside, near the neck of the tube. The location of the deflection unit on the tube is very accurately established by three reference bosses, shown red for clarity. The two 'saddle coils' that form the line coil can be seen in the deflection unit. The frame coil is situated behind, on the plastic carrier. The angular shape of the line coil, the division of the wiring into separate bundles and the permanent magnets above and below the coil are all measures that have been adopted to minimize deflection errors of fifth and higher order (see pp. 170-171).

Fig. 5. Saddle coil for the 20AX (*left*) and 30AX (*right*) systems. For the 20AX the coils have a 'flange' at the neck side, which complicates the structure of the deflection unit, since neither the plastic carrier nor the yoke can be made in one piece. There are no flanges in the 30AX system.

The special features of the new system are thus the greatly reduced manufacturing tolerances and the direct fit between deflection unit and tube, combined with an 'ideal' electron-optical design. The new design is the subject of this article. The most striking of the differences from the 20AX system that result from this new concept will be briefly mentioned here.

Because of the small tolerances in production each tube requires very little correction for *static* convergence (i.e. the convergence when there are no deflection fields). This correction can be effected quite simply by appropriately magnetizing an internal annular magnet from outside the tube, and this is done in the tube factory. With the 20AX system the technician has to correct for static convergence by means of a rather complicated external combination of magnetic multipole cores (see fig. 2).

Since the coils no longer have any 'flanges' on the neck side of the tube (*fig. 5*), the construction can be greatly simplified: the four saddle coils are mounted on a plastic carrier made in one piece, and the Ferroxcube yoke — for short-circuiting the magnetic flux outside the coils — is also in one piece (*fig. 6*).

Owing to the use of slimmer coils, with the copper running close to the neck, and to the longer yoke, the deflection sensitivity is 20% greater. This improves the stability and reliability of operation, since the coils do not become so hot. Since the deflection starts earlier in the neck, the tube had to be made slightly wider at the transition from neck to screen than in the 20AX system, to avoid shadow effects from the tube wall.

Designing a deflection unit requires both calculation and experiment. In the last few years we have developed a software package for computing the deflection fields of deflection units of arbitrary shape (including a Ferroxcube yoke) and for computing the electron trajectories in known deflection fields. In principle it should be possible to design a deflection unit entirely by computer, but in practice this is not quite true. At some stage an experimental model must be made and tested; the results then enable the design to be improved. For the purpose of such experiments we have built equipment for measuring fields and deflection errors.

After a short outline of the essentials of the convergence problem and its solution, we shall first con-

sider the deflection errors that can occur. In addition to the convergence errors mentioned above, these may include 'geometrical' errors, i.e. distortions of the picture. We shall pay particular attention to 'third-order' errors. The elimination of third-order errors would be sufficient if the deflection angle were to remain smaller than about 20°, but there are deflections of 55° in the 20AX and 30AX systems, and errors of up to the 17th order are then perceptible.

Next we shall discuss the mathematical expansion of deflection fields into multipoles. This method enables the errors to be corrected systematically, order by order. It is a distinct improvement on the 'error-coefficient method', which was formerly used and where the deflection fields were expanded as power series of rectangular coordinates [2]. This earlier method is much more complicated, gives less insight into the problem and is limited to errors of the third and fifth order.

These sections on deflection errors and multipoles form an introduction to the treatment of the 'third-order designs' for the 20AX and the 30AX systems; i.e. the designs in which the third-order errors have been eliminated.

The final section deals with the elimination of higher-order errors by means of computer programs and experiments.

The essentials of the convergence problem

We shall now briefly consider the essentials of the convergence problem and its solution in 'in-line' systems.

The horizontal deflection requires a vertically aligned magnetic field, the vertical deflection requires a horizontal field, and it seems only natural to consider homogeneous fields. In a delta system the fields are indeed virtually pure 'dipole fields', i.e. fields that are approximately homogeneous at each cross-section. (We shall define the concept of 'dipole field' more exactly later.) The fields are certainly not homogeneous everywhere; in fact they always vary strongly with the position of the axis: the deflection is concentrated in a limited range. *Fig. 7* gives a diagram of one of the saddle coils and the dipole field excited by the two coils together.

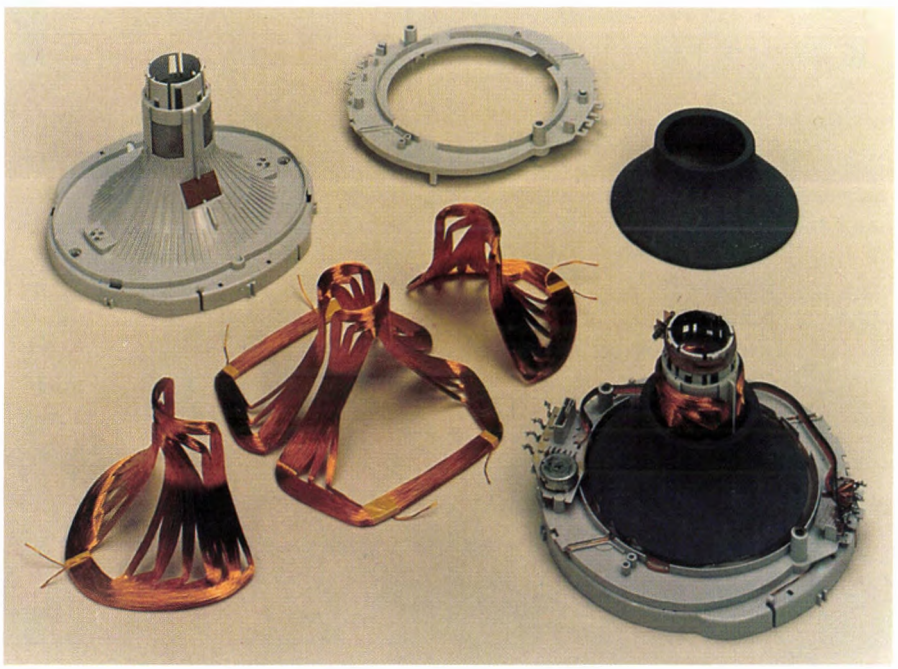


Fig. 6. The 30AX deflection unit (*right foreground*) and the component parts. From *left foreground to right background:* the four saddle coils and the Ferroxcube yoke. *Left background:* the plastic carrier. The 'astigmatism metal' can be seen in the sloping window, and the 'coma metal' in the windows in the neck (see p. 166).

[2] See for example R. Vonk, Philips tech. Rev. **32**, 61, 1971.

We shall now consider first the deflection of *two* beams in such dipole fields. *Fig. 8 (above)* shows the horizontal deflection on the left and the vertical deflection on the right. It can easily be shown (see the caption) that in both cases the point where the beams intersect does not remain on the screen but moves inside the tube.

The remedy for this convergence problem is indicated in fig. 8, *below*. For the horizontal deflection the outer beam must be deflected more strongly than the inner one. The field-strength must therefore increase outwards; this means that the field must be shaped like a 'pin-cushion'. For the vertical deflection the beams must be deflected more towards the outside than in a dipole field; this requires a 'barrel-shaped' field. These corrective measures can be obtained with a four-wire system by appropriately shaping the coils.

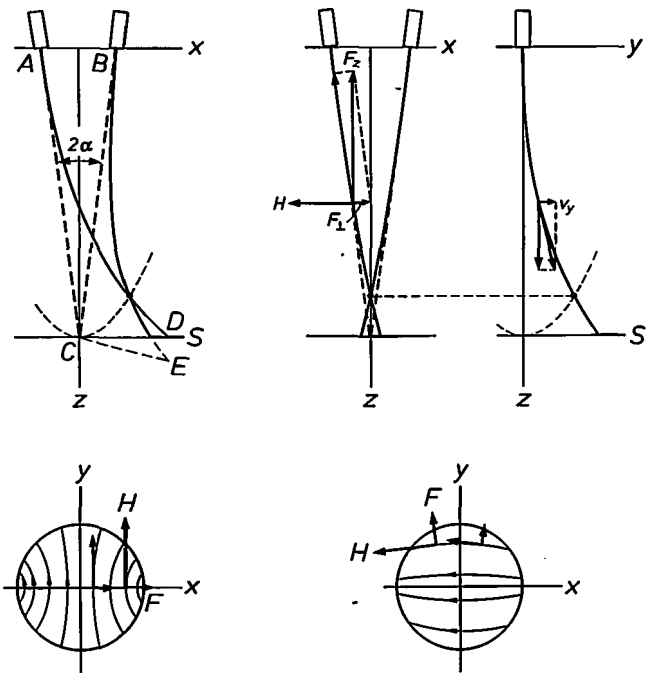


Fig. 8. The convergence problem for two electron beams in dipole fields. *Above, left:* horizontal deflection; *right:* vertical deflection. In both cases the two beams are deflected by practically the same angle in each transverse section. For the horizontal deflection this means that the beam AD is virtually in register with the beam from B when it is rotated about the point C through the angle 2α; D then comes into register with E. From the relative positions of D and E it is immediately obvious that the intersection point of the beams must lie 'behind' the screen S (i.e. inside the tube). In the vertical deflection the vertical component v_y of the electron velocity and the (horizontal) dipole field give a component F_z of the Lorentz force that deflects the path inwards via the component $F_\perp$. Here again the intersection point is displaced inwards. *Below:* the corrective measures; *left:* the 'pin-cushion' field increases outwards in the x-direction, so that the outer beam is deflected more than the inner one; *right:* the barrel-shaped field produces a Lorentz force F directed slightly outwards.

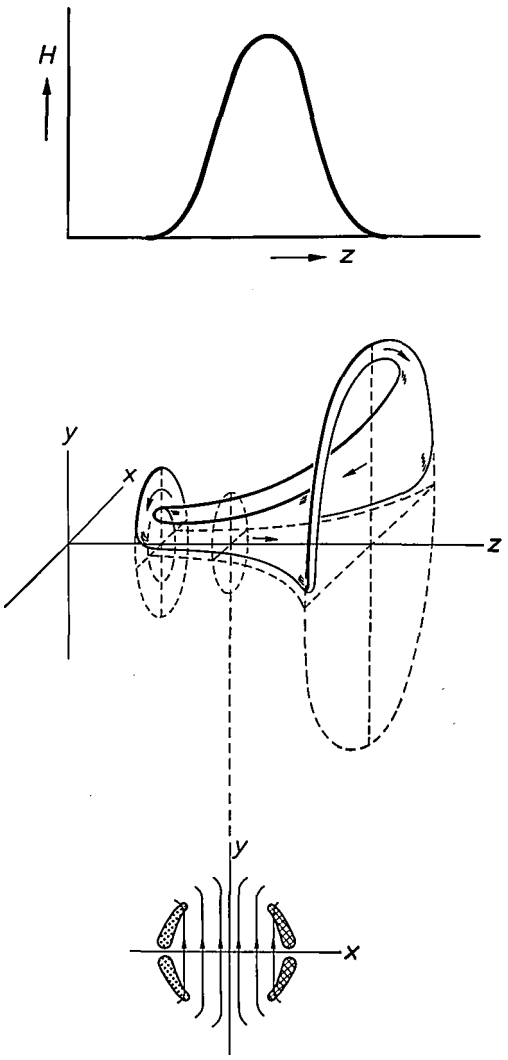


Fig. 7. Diagram of a line saddle coil and the line-coil field for a delta system. The field is virtually a 'dipole field', i.e. roughly homogeneous at every cross-section (*below*). The field-strength, however, is very dependent on the position along the axis (*above*).

It will now be clear what the difficulties with delta systems are. The simple remedy indicated is of no use here because the fields required for each of the three pairs of beams are mutually incompatible. They *are* compatible, however, for in-line systems. In these systems the field of the line coil is indeed a pin-cushion field, and that of the frame coil is a barrel field. Anticipating our discussion of multipoles, we should mention here that such fields can be thought of as composed of 'dipoles' and 'sixpoles' (see fig. 16*a*).

Before we can go into this kind of solution in detail, we should first take a closer look at the possible errors and the multipole expansion of fields.

Deflection errors

We shall now give a brief account of the deflection errors as they appear at the screen.

Because the screen is strongly curved, we project the landing points of the beam on to the plane that is

tangential to the screen at its centre (*fig. 9*); we denote the landing points by the coordinates X, Y in this plane. For small coil currents the deflection in the two directions is proportional to the currents, I_h or I_v. We speak of an 'ideal' deflection when the same proportionalities continue to apply for higher currents. For the coordinates X_0, Y_0 of the 'ideal' landing point P, therefore, we always have:

$$X_0 = aI_h,$$
$$Y_0 = bI_v.$$

In reality, however, the beams R, G and B do not strike the screen exactly at P, and the error $\Delta X, \Delta Y$ is not usually the same for R, G and B. The deviation of the central beam G is called the *geometrical error* ($\Delta X_g, \Delta Y_g$). The convergence errors that the beams R and B have in addition to the geometrical error are known as an *astigmatism error* (half the vector difference of the deviations GR and GB) and *coma* (half the vector sum of GR and GB): see *fig. 10*. Thus, indicating the components of the astigmatism error by $\Delta X_a, \Delta Y_a$ and those of the coma by $\Delta X_c, \Delta Y_c$, then:

$$\begin{aligned} \Delta X_G &= X_G - X_0 = \Delta X_g, \\ \Delta X_R &= X_R - X_0 = \Delta X_g + \Delta X_a + \Delta X_c, \\ \Delta X_B &= X_B - X_0 = \Delta X_g - \Delta X_a + \Delta X_c, \end{aligned} \qquad (1)$$

(and similarly for the ΔY's).

If there is only an astigmatism error and no coma, then R and B lie on opposite sides of G; in the converse situation R and B coincide.

The concepts of astigmatism error and coma can be made clearer by considering a hypothetical 'thick' beam containing R, G and B and focused at the screen in the undeflected state. In an ideal deflection unit for the in-line system, this beam is astigmatic at every deflection, such that a vertical focal line always lies at the screen (the horizontal focal line then lies well behind the screen, i.e. inside the tube). The astigmatism error is associated with an incorrect position of this focal line — not at the screen or not vertical — and the coma is associated with the blurring of this focal line, wherever it is situated.

We assume that the deflection unit has mirror symmetry with respect to the x,z-plane and the y,z-plane, and that the beams in the undeflected state lie in the x,z-plane, G along the z-axis, R and B symmetrically on each side of G (fig. 3). The consequence of these symmetries for the landing points R, G and B on the screen and for the astigmatism error and the coma are shown in *fig. 11*. If the sign of the vertical deflection current I_v is changed, the landing points are reflected in the X-axis. If, however, we change the sign of I_h, then G is reflected in the Y-axis, but R arrives at the mirror image point of B and vice versa, owing to the symmetrical locations of the guns in the x,z-plane.

Third-order deflection errors

We shall now expand the errors $\Delta X_R, \Delta Y_R$ in power series in I_h, I_v and the 'variable' α, which is the angle between R in the undeflected state and the z-axis (see fig. 3). The series for G or B are derived from these by putting α equal to zero or substituting $-\alpha$ for it. A term of the n^{th} degree in I_h, I_v and α together is an 'error of the n^{th} order'.

Since the tube is statically convergent, there are no terms that contain α alone. Nor does the series contain

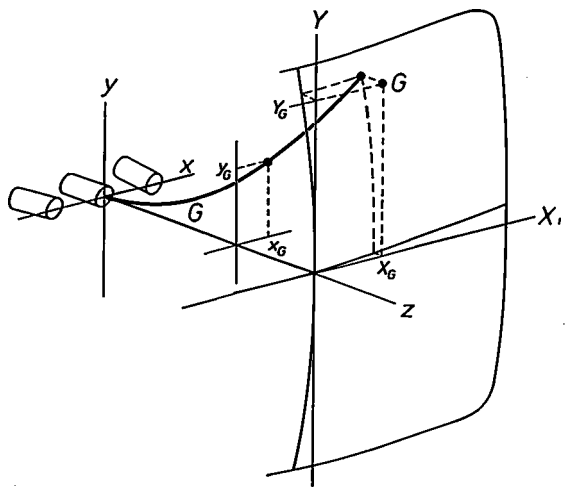


Fig. 9. Coordinate system. The landing points of the beams on the screen are projected on to the X,Y-plane tangential to the screen at its centre. The coordinates x,y of a beam between gun and screen are functions of z (and of the deflection currents). The figure shows these coordinates for the beam G.

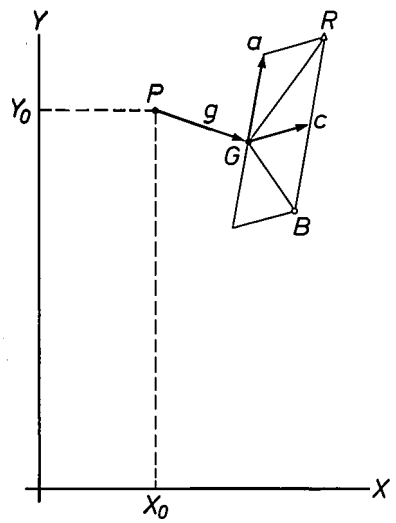


Fig. 10. Deflection errors. If no precautions are taken, the beam G does not arrive at the 'ideal' point P, whose coordinates X_0 and Y_0 are proportional to the magnitudes I_h and I_v of the currents through the line coil and the frame coil. The vector g ($= PG$) is the 'geometrical error'. Also, R and B do not arrive at G. These convergence errors are resolved into an astigmatism error a (half the vector difference of GR and GB) and a coma c (half the vector sum).

first-degree terms in I_h and I_v, since these represent the ideal deflections themselves. Because of the symmetries discussed above, many other terms are also missing, including *all terms of even degree* (even in α, I_h and I_v together). If we also neglect all terms of the fifth degree and higher, we only retain terms of the third order as possible deflection errors. These are listed in *Table I*. The geometrical error given in Table I is split into the *'S-distortion'* (or *linearity error*), which stretches the picture at the edges in the horizontal and vertical directions, and the *EW*- and *NS*-distortions, which introduce pin-cushion or barrel distortion into the picture (*fig. 12*).

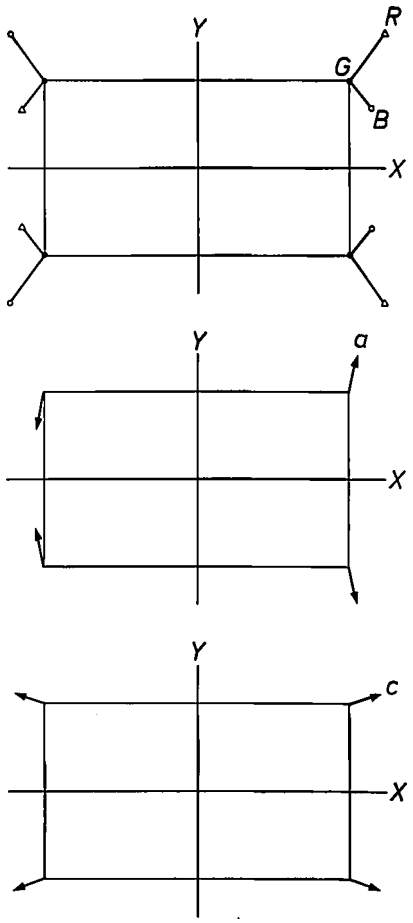


Fig. 11. Symmetry properties of points R, G and B on the screen (*above*), of the astigmatism error (*centre*) and of the coma (*below*) when I_h or I_v or both change sign. The two lower figures are derived from the upper one (see fig. 10). The symmetry depends on the symmetry of the deflection unit and of the arrangement of the electron guns.

From the symmetry that is found when I_h changes sign — reflection in the Y-axis with R and B interchanged — we may conclude, for example, that:

$$\Delta X(-I_h,I_v,-\alpha) = -\Delta X(I_h,I_v,\alpha),$$
$$\Delta Y(-I_h,I_v,-\alpha) = \Delta Y(I_h,I_v,\alpha),$$

and from this it follows that the series expansion of ΔX contains no terms that are of *even* degree in I_h and α together (e.g. $I_h^2 I_v$, $I_h I_v \alpha$, $I_v \alpha^2$), and that the expansion of ΔY contains no terms that are of *odd* degree in I_h and α together. When all the consequences of the symmetries are similarly taken into account, and all terms of fifth and higher degree are neglected, only the following terms are left:

$$\Delta X_R = a_3 I_h^3 + a_1 I_h I_v^2 + (b_2 I_h^2 + b_0 I_v^2)\alpha + c_1 I_h \alpha^2,$$
$$\Delta Y_R = a_0 I_v^3 + a_2 I_h^2 I_v + b_1 I_h I_v \alpha + c_0 I_v \alpha^2.$$

As we said earlier, the errors for G or B follow from this when 0 or $-\alpha$ are substituted for α. By comparing this result with (1) we find the various contributions listed in Table I. The terms of zero degree in α are geometrical errors, those of first degree in α are astigmatism errors, and those of second degree in α are associated with coma.

Table I. The possible third-order errors of a symmetrical deflection unit.

Geometrical error	$\Delta X_g = \Delta X_S + \Delta X_{EW}$ $\Delta Y_g = \Delta Y_S + \Delta Y_{NS}$
S-distortion	$\Delta X_S = a_3 I_h^3$ $\Delta Y_S = a_0 I_v^3$
EW-distortion	$\Delta X_{EW} = a_1 I_h I_v^2$
NS-distortion	$\Delta Y_{NS} = a_2 I_h^2 I_v$
Astigmatism error	$\Delta X_a = (b_2 I_h^2 + b_0 I_v^2)\alpha$ $\Delta Y_a = b_1 I_h I_v \alpha$
Coma	$\Delta X_c = c_1 I_h \alpha^2$ $\Delta Y_c = c_0 I_v \alpha^2$

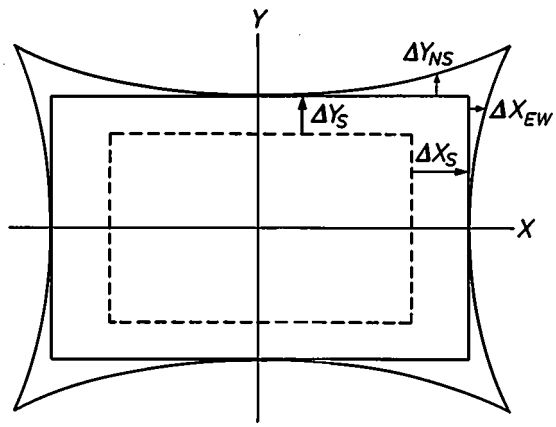


Fig. 12. Geometrical errors of the third order. The dashed 'ideal' rectangle undergoes S-distortion (ΔX_S,ΔY_S), EW-distortion (ΔX_{EW}) and NS-distortion (ΔY_{NS}). On account of the S-distortion, X is no longer proportional to I_h nor Y to I_v, but the figure remains rectangular. The S-distortion is usually compensated by making the currents vary with time in an S-curve rather than linearly. In the diagram ΔX_{EW} and ΔY_{NS} are positive, and the distortion is of 'pin-cushion' type; when ΔX or ΔY are negative the distortion is 'barrel distortion'.

Multipoles

The procedure for the multipole expansion of deflection fields is as follows. We choose a set of cylindrical coordinates (r,ψ,z) in which the z-axis coincides with the coil axis (*fig. 13*), and expand the

field in the deflection space into a Fourier series in the angular variable ψ. Each term in this series forms a multipole component of the deflection field.

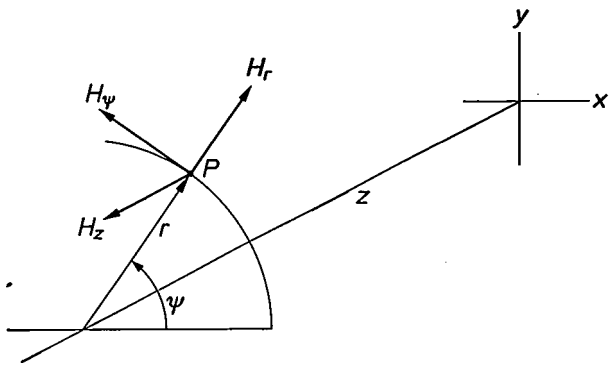


Fig. 13. The cylindrical coordinate system (r, ψ, z) used in the multipole expansion, and the three components H_r, H_ψ and H_z of the magnetic field in this system. The z-axis is taken along the axis of the deflection unit.

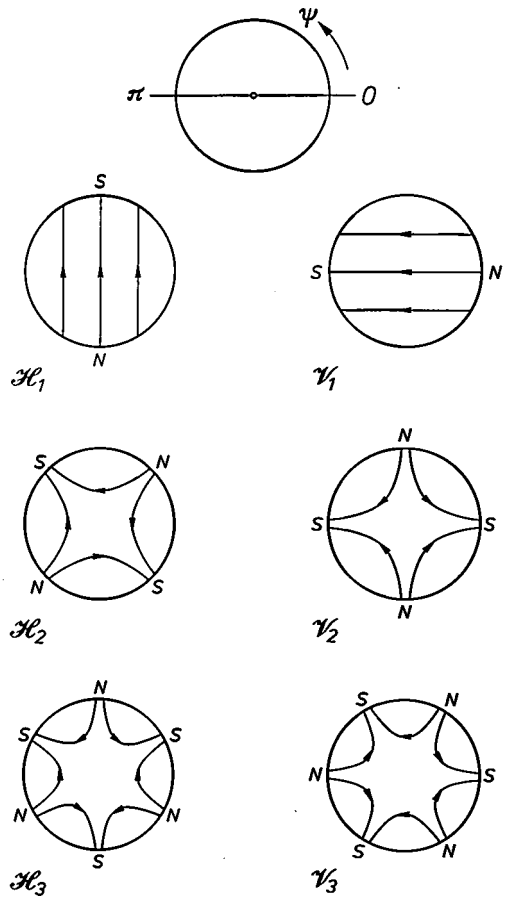


Fig. 14. Multipole fields in the r, ψ-plane. The multipoles are divided into 'horizontal' or 'line' multipoles ($\mathcal{H}_1$, $\mathcal{H}_2$, ...) and into 'vertical' or 'frame' multipoles ($\mathcal{V}_1$, $\mathcal{V}_2$, ...). They are described by the 'magnetic potentials' $r\sin\psi$, $r^2\sin 2\psi$, ... and $r\cos\psi$, $r^2\cos 2\psi$, ... respectively (see eq. (4)). From this it follows (see eq. (3)) that the components H_r and H_ψ in $\mathcal{H}_1$ and $\mathcal{V}_1$ are independent of r, while the components in $\mathcal{H}_2$ and $\mathcal{V}_2$ are proportional to r, the components in $\mathcal{H}_3$ and $\mathcal{V}_3$ to r^2, and so on. The line multipoles are antisymmetric, the frame multipoles are symmetric with respect to the axis $\psi = 0$.

The field variation in the r, ψ-plane for the simplest multipoles found in this way, the 'line' and 'frame' dipoles, fourpoles and sixpoles, is shown in *fig. 14*. The number of north poles (or south poles or periods) n is the 'multiplicity' of a $2n$-pole ($\mathcal{H}_n$ or $\mathcal{V}_n$). The dipole $\mathcal{H}_1$ is the principal component of the line-coil field and the dipole $\mathcal{V}_1$ is the principal component of the frame-coil field. The amplitude of each multipole is a function of z; in designing a deflection unit each multipole must be given the correct amplitude function.

It follows from the NS-symmetry of the coils that the field of the line coil is NS-*anti*symmetric, and the field of the frame coil NS-symmetric (*fig. 15*). Comparison with fig. 14 shows that the line coil can therefore only generate 'line multipoles', and the frame coil only 'frame multipoles' (hence these names). It also follows from the EW-symmetry of the coils that the field of the line coil is EW-symmetric and the field of the frame coil is EW-antisymmetric; from this it follows that no fourpoles are excited in either case, nor any other multipoles with even multiplicity. Only dipoles, sixpoles, tenpoles, and so on will be excited.

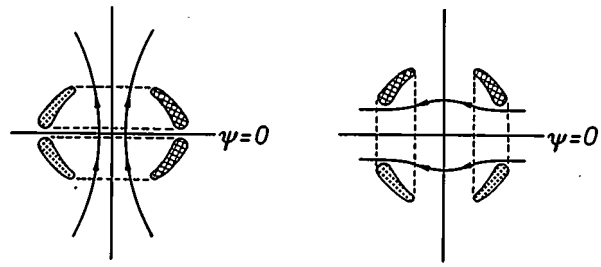


Fig. 15. Symmetries of the deflection fields. *Left:* a cross-section through the line coil; *right:* a cross-section through the frame coil. *Hatched areas:* the current flows into the paper; *points:* the current flows out of the paper. From the NS and EW symmetry of the structure it follows that the field for the line coil is NS-*anti*symmetric and EW-symmetric. The opposite applies for the frame coil. The figures represent instantaneous values of the current and field.

To illustrate the analysis of arbitrary (two-dimensional) fields into multipoles *fig. 16* shows how a dipole and a sixpole produce a pin-cushion field, and how current imbalance produces a fourpole. Conversely, disturbed symmetry can be restored with the aid of fourpoles.

In the correction of deflection errors the multipole method offers two major advantages. In the first place a deflection unit can be designed in steps, because *a multipole has no effect on deflection errors of lower order than its multiplicity.* The principal action of a $2n$-pole is its influence on n^{th}-order errors; all subsidiary effects are always of higher order. Pure dipoles are therefore used for the deflection, and the resultant

third-order errors are corrected with sixpoles. This is as far as we shall go in the next section, but we could in principle continue in the same way, correcting the remaining fifth-order errors with tenpoles, and so on.

In the second place there is a simple relation, in principle at least, between a multipole distribution and the wire distribution required for it: if the density of the current perpendicular to the r,ψ-plane is also expanded into a Fourier series in ψ, each harmonic excites the corresponding multipole. As an illustration

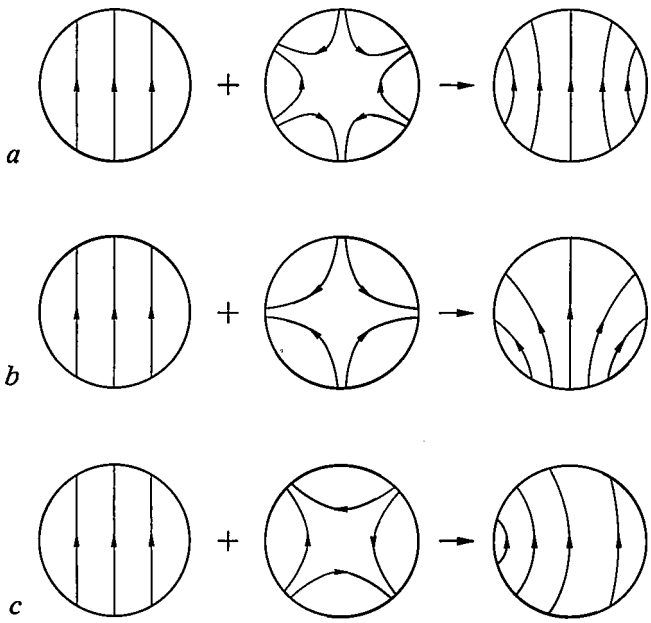


Fig. 16. Examples of the composition of multipoles to form other fields. *a*) A dipole and a sixpole produce a pin-cushion field. A barrel-type field is obtained with a sixpole of opposite sign (*N* and *S* reversed). *b*) and *c*) A dipole and a fourpole produce an asymmetric field. Conversely, fourpoles can be introduced to restore disturbed symmetry. If the currents in the two saddle coils forming the line coil are not equal, there is asymmetry of type (*b*).

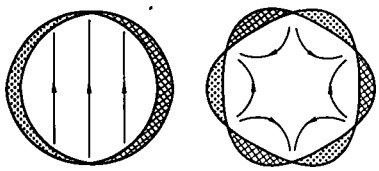


Fig. 17. Current-density distributions for generating a pure dipole and a pure sixpole. The current density varies sinusoidally along the circumference.

fig. 17 shows the wire distribution for generating a pure dipole and a pure sixpole. To produce a $2n$-pole with a given coil we must therefore alternately concentrate and 'thin out' the wiring along the circumference, with a periodicity of $2\pi/n$.

At each point of the deflection space the field-strength of a line multipole varies in proportion to I_h, and the field-strength of the frame multipole varies in proportion to I_v. This is still true if the field is distorted with soft magnetic ('linear') material, e.g. the Ferroxcube of the yoke. An effect on the electron beams can also be produced with *permanent* magnets. In their effect on the beams, there is a simple relation between the 'static' multipoles thus produced and the 'dynamic' coil multipoles: a static $2(n+1)$-pole is equivalent to the sum of a dynamic line $2n$-pole and a dynamic frame $2n$-pole.

The mathematical treatment of multipoles can be briefly summarized by making use of the scalar magnetic potential Φ of the field in the deflection space. This potential must satisfy the Laplace equation:

$$\Delta\Phi = 0, \tag{2}$$

and the field is derived from Φ by means of the relations

$$H_z = -\frac{\partial\Phi}{\partial z},\quad H_\psi = -\frac{1}{r}\frac{\partial\Phi}{\partial\psi},\quad H_r = -\frac{\partial\Phi}{\partial r}. \tag{3}$$

The multipole expansion is then:

$$\Phi = \sum_{n=1}^{\infty} A_n r^n \sin n\psi + B_n r^n \cos n\psi. \tag{4}$$

The sine terms represent the line multipoles, the cosine terms the frame multipoles. For symmetrical coils there are no terms in which n is even. With constants for A_n and B_n expression (4) satisfies equation (2). The field is then 'two-dimensional': it does not depend on z. For 'three-dimensional' fields the A_n's and the B_n's are functions of r and z, and differential equations for these functions can be derived from eq. (2). As soon as the form of the A_n's and the B_n's on a single curve in the r,z-plane is known, e.g. on the axis or on the contour of the deflection space, they can be calculated from these equations in the entire r,z-plane; we then know the field in the deflection space as well.

The relation between static and dynamic multipoles is derived from the fact that the field-strength (or the potential) of a dynamic multipole at a particular point of a particular cross-section is always the same when the beam passes that point. Thus, at a point with the horizontal coordinate $x(=r\cos\psi)$, $\mathcal{H}_n$ is equivalent to a static potential $x \times r^n \sin n\psi = r^{n+1} \sin n\psi \cos\psi$, because the magnitude of the line-coil current I_h is always equal to x, apart from a constant factor, when the beam passes that point. Similarly, $\mathcal{V}_n$ is equivalent to a static potential $r^{n+1}\cos n\psi \sin\psi$. The sum of both, $r^{n+1}\sin(n+1)\psi$, is the potential of a static $(n+1)$-pole. Strictly speaking, this argument is only applicable to 'two-dimensional' fields.

Third-order designs

As we have already seen in fig. 8, pure dipoles give a severe horizontal astigmatism error, both for horizontal and for vertical deflection; the S-, EW- and NS-distortions are also fairly pronounced. For a deflection of 55° on a screen of 66 cm diagonal, calculation of the electron paths gives about 30 mm for the horizontal astigmatism error, 40 mm for the S-distortion and 25 mm for the EW- and NS-distortions. The vertical astigmatism error, the coma and the higher-order deflection errors each amount to no more than a few millimetres. The convergence errors are shown in

fig. 18. We shall now look first at the principal action of sixpoles, and then we shall discuss the reduction of the errors by means of sixpoles.

The principal action of sixpoles

When the central beam G of the three beams passes through the centre of a sixpole or other multipole, the effect can immediately be derived from the pattern of the lines of force and the resultant Lorentz forces (*fig. 19*). For example, $\mathcal{H}_2$ exerts horizontal forces on R and B that are symmetrical with respect to G; this fourpole therefore makes a contribution to the horizontal astigmatism error. The contribution is 'positive' because R is subjected to a force towards the right. The results for two-, four- and sixpoles are collected in *Table II*. The effect of the fourpoles is proportional to the *eccentricity* e (the distance from R to G at the location of the fourpole), because the field of a fourpole is proportional to r (see fig. 14); similarly for a sixpole the effect is proportional to e^2. Moreover, each local transverse force acting on the beam is converted into a displacement on the screen that is proportional to the distance d from the point of action to the screen. Since e is proportional to d, the 'sensitivities' of the errors for dipoles, fourpoles and sixpoles are thus in total proportional to d, d^2 and d^3. Table II can be used for multipoles in the neck of the tube, where the deflection of the electron paths is still very small.

If the beam G does *not* pass through the centre of the multipole, we expand the multipole into a series of multipoles centred on G and consult Table II for each. For sixpoles this expansion is

$$\mathcal{H}_3 = \mathcal{H}_3^G + 3[(x_G^2 - y_G^2)\mathcal{H}_1^G - 2x_G y_G \mathcal{V}_1^G - x_G \mathcal{H}_2^G + y_G \mathcal{V}_2^G], \quad (5)$$

$$\mathcal{V}_3 = \mathcal{V}_3^G + 3[-2x_G y_G \mathcal{H}_1^G + (-x_G^2 + y_G^2)\mathcal{V}_1^G + y_G \mathcal{H}_2^G + x_G \mathcal{V}_2^G], \quad (6)$$

where $x_G(z)$ and $y_G(z)$ are the coordinates of G in the transverse section considered (see fig. 9). The $\mathcal{H}^G$'s are multipoles centred on the beam G and with line-coil orientation; the $\mathcal{V}^G$'s are similar but with frame-coil orientation.

We now consider the effect of $\mathcal{H}_3$. The magnitudes of $\mathcal{H}_3^G, \mathcal{H}_1^G, \ldots$ in (5) are proportional to that of $\mathcal{H}_3$ and hence to I_h, since $\mathcal{H}_3$ originates from the line coil. The principal term $\mathcal{H}_3^G$ thus makes a contribution to the coma (see Table II) that is proportional to $d^3 I_h$.

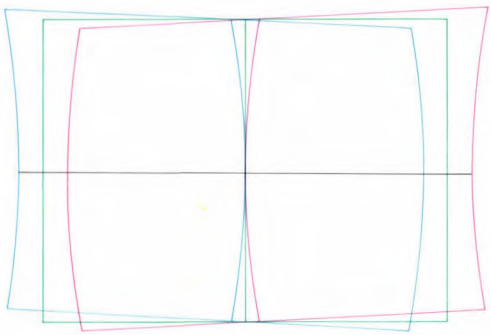


Fig. 18. Convergence errors of pure dipoles that produce the desired deflection (55°). The horizontal astigmatism error is dominant. The errors are magnified by ten times compared with the picture dimensions. The geometrical error is not taken into account here.

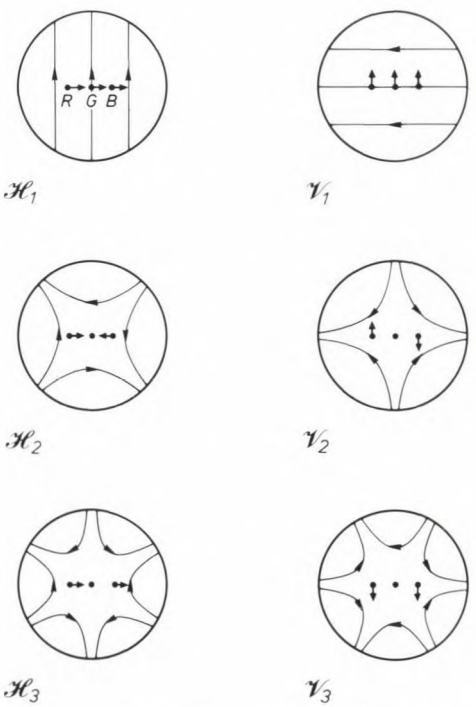


Fig. 19. Effect of multipoles on the three beams when the central beam G passes through the centre of the multipole. The multipoles are seen from the front end (the screen), so that the beam R lies on the left (see fig. 3). The velocity vector of the electrons is in the forward direction coming out of the plane of the paper. The Lorentz forces resulting from the field patterns are indicated by the bold arrows. The resultant contributions to the deflection errors are listed in Table II.

Table II. The principal action of two-, four- and sixpoles when the central beam G is concentric. In the last column, d is the distance of the (local) multipole from the screen; e is the eccentricity.

			Sign	Sensitivity
$\mathcal{H}_1$	horiz. deflection	ΔX_g	+	d
$\mathcal{V}_1$	vert. deflection	ΔY_g	+	d
$\mathcal{H}_2$	horiz. astig. error	ΔX_a	+	$ed \propto d^2$
$\mathcal{V}_2$	vert. astig. error	ΔY_a	+	$ed \propto d^2$
$\mathcal{H}_3$	horiz. coma	ΔX_c	+	$e^2 d \propto d^3$
$\mathcal{V}_3$	vert. coma	ΔY_c	−	$e^2 d \propto d^3$

For x_G and y_G we have the approximate expressions

$$x_G = f(z)I_h,\ y_G = f(z)I_v,$$

where $f(z)$ gives the shape of the electron trajectory in a dipole field. The first term in the square brackets, $(x_G^2 - y_G^2)\mathcal{H}_1^G$, therefore makes a contribution ΔX_g to the horizontal deflection (see $\mathcal{H}_1$ in Table II) that is proportional to $df^2(I_h^2 - I_v^2)I_h$, and therefore makes a positive contribution to the horizontal S-distortion ($df^2I_h^3$) and a negative one to the EW-distortion ($-df^2I_hI_v^2$). The term $-2x_Gy_G\mathcal{V}_1^G$ makes a contribution ΔY_g to the vertical deflection (see $\mathcal{V}_1$ in Table II) that is proportional to $-2df^2I_h^2I_v$ (negative NS-distortion). In a similar way we find for the following two terms contributions to the horizontal astigmatism error ($-d^2fI_h^2$) and the vertical astigmatism error ($d^2fI_hI_v$).

These effects are collected in *Table III*, together with the effects of $\mathcal{V}_3$ and the third-order errors of the dipoles. The factors d (the distance to the screen) and f determine where an effect is most sensitive for the sixpole. Since f increases (nonlinearly) from neck to screen the effect of the sixpoles on the geometrical errors is by far the greatest on the screen side of the coil (df^2), whereas its effect on the coma is greatest on the neck side (d^3); the effect on the astigmatism error is greatest from the screen side to the centre (d^2f). The factors mentioned here are only indications; an accurate calculation leads to the sensitivity curves of *fig. 20*.

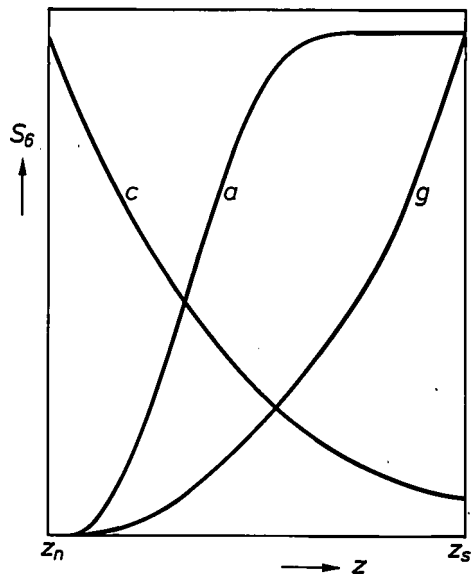


Fig. 20. The sensitivity S_6 of the geometrical error (g), the astigmatism error (a) and the coma (c) for sixpoles, as a function of the position z of the sixpole on the axis; z_n is the end of the coil on the neck side, z_s is the end of the coil on the screen side. Sixpoles for correcting coma should preferably be applied to the neck side, and those for correcting the geometrical error should preferably be near the screen; the astigmatism error can also be effectively dealt with at the centre.

Table III. Contributions to third-order errors of dipoles and sixpoles.

		Dipoles		$\mathcal{H}_3$	$\mathcal{V}_3$
horiz. S-distn	ΔX_S	$a_3I_h^3$	(+)	$df^2I_h^3$	
vert. S-distn	ΔY_S	$a_0I_v^3$	(+)		$df^2I_v^3$
EW-distn	ΔX_{EW}	$a_1I_hI_v^2$	(+)	$-df^2I_hI_v^2$	$-2df^2I_hI_v^2$
NS-distn	ΔY_{NS}	$a_2I_h^2I_v$	(+)	$-2df^2I_h^2I_v$	$-df^2I_h^2I_v$
horiz. astig. error	ΔX_a	$(b_2I_h^2+b_0I_v^2)\alpha$	(+)	$-d^2fI_h^2$	$d^2fI_v^2$
vert. astig. error	ΔY_a	$(b_1I_hI_v\alpha)$		$d^2fI_hI_v$	$d^2fI_vI_h$
horiz. coma	ΔX_c	$(c_1I_h\alpha^2)$		d^3I_h	
vert. coma	ΔY_c	$(c_0I_v\alpha^2)$			$-d^3I_v$

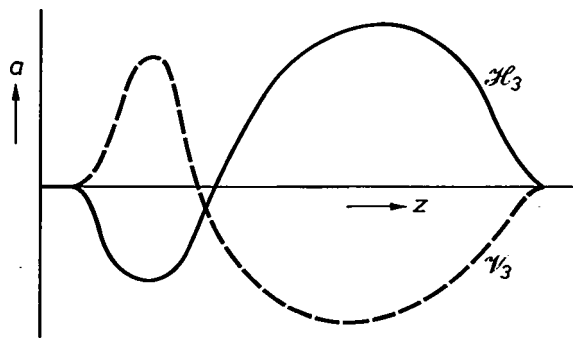


Fig. 21. The amplitude functions $a(z)$ of $\mathcal{H}_3$ and $\mathcal{V}_3$ for the 20AX system.

The third-order design for the 20AX system

Since the coefficients b_0 and b_2 for the horizontal astigmatism error of the dipoles are not very different, Table III and fig. 20 show that this error can be corrected with a positive $\mathcal{H}_3$ and a negative $\mathcal{V}_3$ of about the same magnitude, somewhere between the screen side and the centre. This gives rise to other effects, however. To a first approximation, the contributions to the vertical astigmatism error cancel out, because of the opposite sign of the sixpoles. The coma produced by the sixpoles at the centre has to be corrected, however, and this can be done by means of an opposing sixpole near the neck. In this way we arrive at the amplitude functions in *fig. 21* for the 20AX. In the line coil the wiring near the screen must therefore be more concentrated around the horizontal plane than the wiring for pure dipoles, and near the neck the wiring must be more concentrated around a plane 60° out of the horizontal; see *fig. 22*. The corresponding measures for the frame coil are less effective, partly because the wiring is further away from the axis; the required positive value of $\mathcal{V}_3$ near the neck cannot be completely attained in this way.

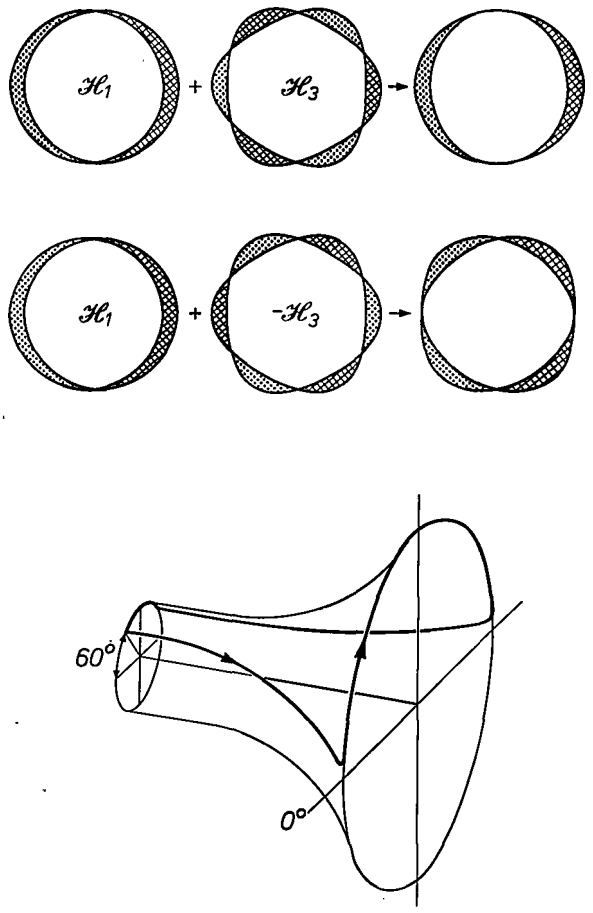


Fig. 22. Wire distribution required for the amplitude function of $\mathscr{H}_3$ in fig. 21. *Top row:* combination of the wiring for an $\mathscr{H}_1$ with that for a positive $\mathscr{H}_3$ results in wiring that is more concentrated around the horizontal plane than for an $\mathscr{H}_1$. *Lower row:* to introduce a negative $\mathscr{H}_3$ the wiring should be more concentrated around a plane at an angle of 60° to the horizontal plane. *Bottom:* wiring which is at 0° near the screen and at 60° near the neck. This pattern can be distinguished in the various photographs.

The additional measures taken to achieve this in the 20AX consist of taking 'bites' out of the Ferroxcube yoke on the neck side.

The effect of the sixpoles on the geometrical errors is mainly determined by the value near the screen. In the 20AX the NS-distortion of the (positive) $\mathscr{H}_3$ roughly compensates the distortion of the dipoles and of the (negative) $\mathscr{V}_3$. Since the coefficients a in Table III are not very different either, the vertical S-distortion is then also roughly cancelled out; the EW-distortion and the horizontal S-distortion are enhanced, however.

Although the contributions of the $\mathscr{H}_3$ and the $\mathscr{V}_3$ to the vertical astigmatism error roughly cancel out, the remaining error, including the contributions from the dipoles, is still too large; it is negative and is approximately 3 mm from R to B in a 66 cm tube. Hence the asymmetry of $\mathscr{H}_3$ and $\mathscr{V}_3$ in fig. 21: $\mathscr{V}_3$ is made less negative to an amount sufficient to correct the vertical astigmatism error. This in turn gives rise to a horizontal astigmatism error, however. Table II shows that this can be corrected with an $\mathscr{H}_2$, which must not however originate from the deflection coils, because its magnitude must be proportional to I_V^2. This fourpole is excited by four toroidal windings around the yoke. By adding other components to the supply current, this fourpole can also be used for eliminating possible symmetry errors; such errors are not unusual with the 20AX system owing to the relatively large tolerances in production.

The third-order design for the 30AX system

In designing the deflection unit for the 30AX system, one of the objectives was to remove the severe EW-distortion occurring in the 20AX system, while maintaining complete convergence.

Table III shows that this distortion can be reduced by making $\mathscr{V}_3$ less strongly negative near the screen. This would of course give an astigmatism error again, but this can be prevented by making $\mathscr{V}_3$ more negative at the centre. The coma that then appears can be compensated by making $\mathscr{V}_3$ more positive near the neck. For the 30AX frame coil we thus arrive at an amplitude function of $\mathscr{V}_3$ as illustrated in *fig. 23*.

Now a function like this cannot be obtained by using the conventional winding technique. The wiring near the neck would have to be concentrated strongly around the vertical plane, and the wiring near the centre would have to be concentrated strongly around a plane at an angle of 60° to the vertical, and finally the wiring near the screen would again have to be concentrated near the vertical, but now fairly weakly — somewhat more weakly than for a pure dipole

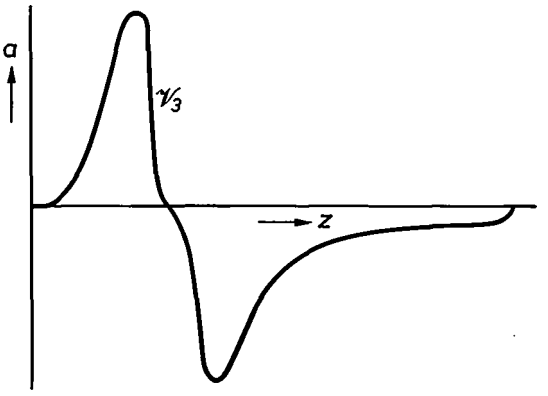


Fig. 23. The amplitude function $a(z)$ of $\mathscr{V}_3$ for the 30AX.

(*fig. 24*). The difficulties with such a coil become clear if we take a closer look at the winding technique as illustrated by *fig. 25*. This shows a coil that *can* be wound — a line saddle coil — on one half of the winding jig. The other half of the jig mates with this, leaving the winding gap; the coil is wound in this gap

from inside to outside. It will be clear that the turns wound in this way can be made to bulge outward by putting obstacles in the gap, but that it is not possible to make *indentations*, although this is exactly what fig. 24 shows would be necessary.

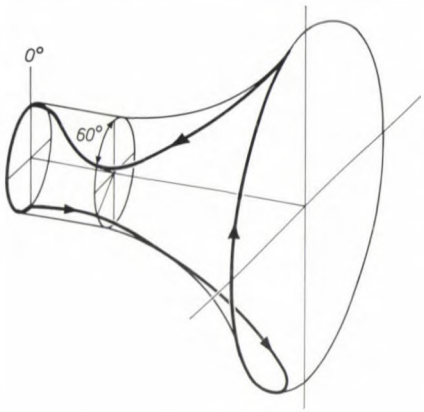


Fig. 24. Wiring pattern that would be required for the amplitude function in fig. 23: a strong concentration around the vertical near the neck, a strong concentration around a plane at an angle of 60° to the vertical at the centre, and again a concentration around the vertical, but much weaker, near the screen. The indentation in the pattern cannot be produced with the winding technique now employed.

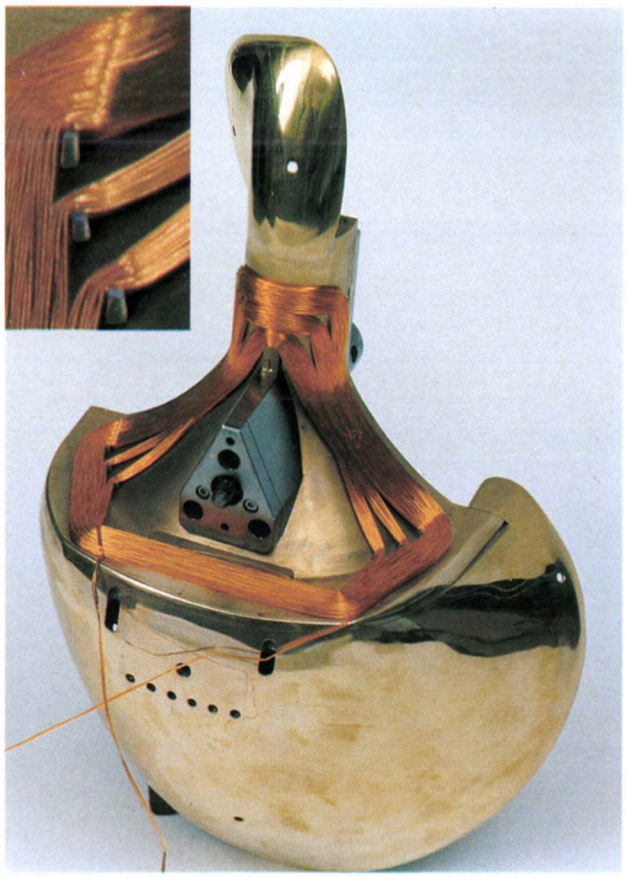


Fig. 25. Line saddle coil on one half of a winding jig. When the two halves are put together a 'gap' is formed; the coil is wound in this gap. The wire is fed in by means of a rotary arm and guided into the gap by the shiny 'helmet' at the bottom. *Inset:* the required shape is obtained with the aid of pins that are 'shot' into the gap during the winding process.

We have solved this problem by using soft-magnetic material, which we call 'astigmatism metal' and 'coma metal' (see fig. 6). *Fig. 26* shows how the 'primary' dipole field magnetizes these metal parts to produce a secondary field that contains a strong negative or positive sixpole component. The 'astigmatism metal' thus produces the desired negative Y_3 peak at the centre, and the 'coma metal' provides the positive peak near the neck. The weaker Y_3 component near the screen can be obtained by using an appropriate distribution for the wiring.

After the coarse errors have been eliminated in this way, the remaining third-order errors have to be dealt with. These are the vertical astigmatism error and a slight NS-distortion, which has crept in because Y_3

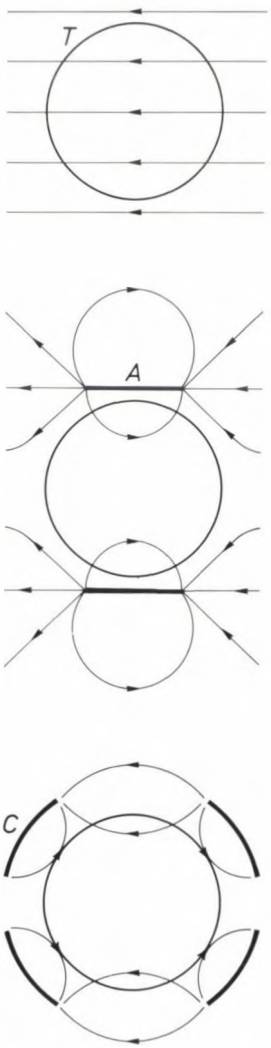


Fig. 26. Application of soft-magnetic material in the 30AX system. *Above:* the primary dipole field; *T* cross-section of the tube. *Centre:* the secondary field of the astigmatism metal (*A*) magnetized by the primary field. *Below:* as before, but now for the coma metal (*C*). The astigmatism metal generates in the first instance a negative Y_3 (see. fig. 14); it is mounted in the centre of the deflection unit and produces the negative peak in fig. 23. The coma metal excites a positive Y_3; it is mounted near the neck giving the positive peak in fig. 23. Both arrangements can be seen in fig. 6.

has been made less negative near the screen. By paying closer attention to the possibilities of varying the wire distribution, we have succeeded in removing all third-order errors simultaneously in the 30AX system, without using coils that require a separate supply of current, such as the fourpole toroid in the 20AX system.

The method used here consists briefly in displacing the path of the beam G while retaining the same landing point on the screen, in other words in changing the functions $x_G(z)$ and $y_G(z)$ between gun and screen (see fig. 9). If for example the frame coil is lengthened on the neck side, the vertical deflection starts earlier and changes $y_G(z)$. The function $x_G(z)$ can be altered independently of $y_G(z)$ by modifying the shape of the line coil. To illustrate the scope that this offers, we have modified Table III slightly by replacing the term $df^2I_h^3$ with the original expression $dx_G^2I_h$ and so on; the result is *Table IV*. It can be seen that the effect of x_G and y_G variations on the individual deflection errors differs widely. Consequently we were able to eliminate all the smaller third-order errors.

Table IV. Contribution to deflection errors of sixpoles.

	$\mathcal{H}_3$	$\mathcal{V}_3$
horiz. S-distn	$dx_G^2I_h$	
vert. S-distn		$dy_G^2I_v$
EW-distn	$-dy_G^2I_h$	$-2dx_Gy_GI_v$
NS-distn	$-2dx_Gy_GI_h$	$-dx_G^2I_v$
horiz. astig. error	$-d^2x_GI_h$	$d^2y_GI_v$
vert. astig. error	$+d^2y_GI_h$	$d^2x_GI_v$
horiz. coma	$+d^3I_h$	
vert. coma		$-d^3I_v$

Reducing deflection errors of higher order

In this section we shall first briefly review the experimental equipment and computer software that we have developed in the last few years for eliminating higher-order errors, and we shall then give an actual example of a fifth-order error to show how these aids have been used in the development of the 30AX coils.

In brief, the experimental work amounts to making an experimental model at a certain stage in the design, measuring both the deflection errors and the multipole distributions of the fields, and looking for a clear relation between the two groups of measured results. Starting from such a relation, we then try to improve the multipoles by an appropriate distribution of wire and magnetic material. If this attempt fails because a clear relation cannot be established, we can resort to computer-optimization programs.

[3] The field-measuring equipment will be the subject of a forthcoming article in this journal.

Deflection errors and multipole composition

To measure the deflection errors we mount the experimental model on a typical tube. A video camera is used to record the deflection errors in a network of points on the screen. The results are processed to reproduce graphs of the errors as a function of the location on the screen. Both the positioning of the camera and the processing are fully automatic. The accuracy to which the convergence is measured is about 0.05 mm, and for the geometrical error the accuracy is about 0.5 mm.

To determine the multipole composition of a field the best course is to measure the component H_ψ on a series of circles around the axis. When H_ψ on such a circle is expanded in a Fourier series in ψ, a simple calculation applied to each harmonic gives the amplitude of the associated multipole at that particular location. The determination becomes more accurate as the circles increase in size; each circle should therefore be about as large as the tube cross-section at that position. *Fig. 27* is a diagram of the equipment used for the measurements [3]. *Fig. 28* gives a typical result.

The relation between the Fourier components of H_ψ on a circle around the axis and the local multipole amplitudes follows from the Fourier expansion of H_ψ found by substituting (4) in the expression for H_ψ in (3). The result is:

$$H_\psi = -\sum_{n=1}^{\infty} nA_nr^{n-1}\cos n\psi + nB_nr^{n-1}\sin n\psi.$$

Thus, from each Fourier amplitude of H_ψ and the radius r of the circle an amplitude A_n or B_n can be derived directly. Except for $n = 1$, the Fourier terms of H_ψ increase with r and can therefore be determined more accurately as the value of r increases.

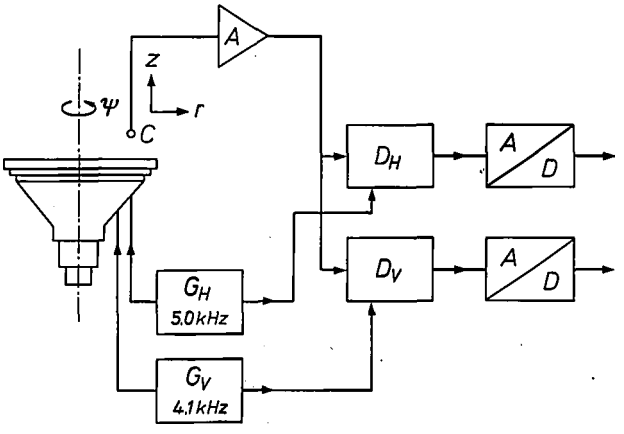


Fig. 27. Diagram of the field-measuring equipment. The ψ-component of the field of the deflection unit is measured by the coil C at a number of points on a series of circles; a new measurement point is obtained by rotating the unit, and a new circle by displacing C in the r,z-plane. The two deflection fields can be analysed simultaneously by feeding sinusoidal signals of different frequency (G_H,G_V), to the line coil and the frame coil and synchronously detecting the measured signals (D_H,D_V). After digital conversion (A/D) both signals are processed by a P855 minicomputer, which analyses the measured fields into multipoles.

A relation must now be established between the measured errors and the multipole composition derived. After some experience it is possible to recognize all kinds of errors at first sight. *Fig. 29*, for example, shows some familiar NS-distortions of

If a single characteristic error is now dominant, it is usually clear what is wrong with the multipoles and an improved design can be drawn up on paper. Once the obvious errors have been eliminated, any remaining error can be reduced by using a computer optimi-

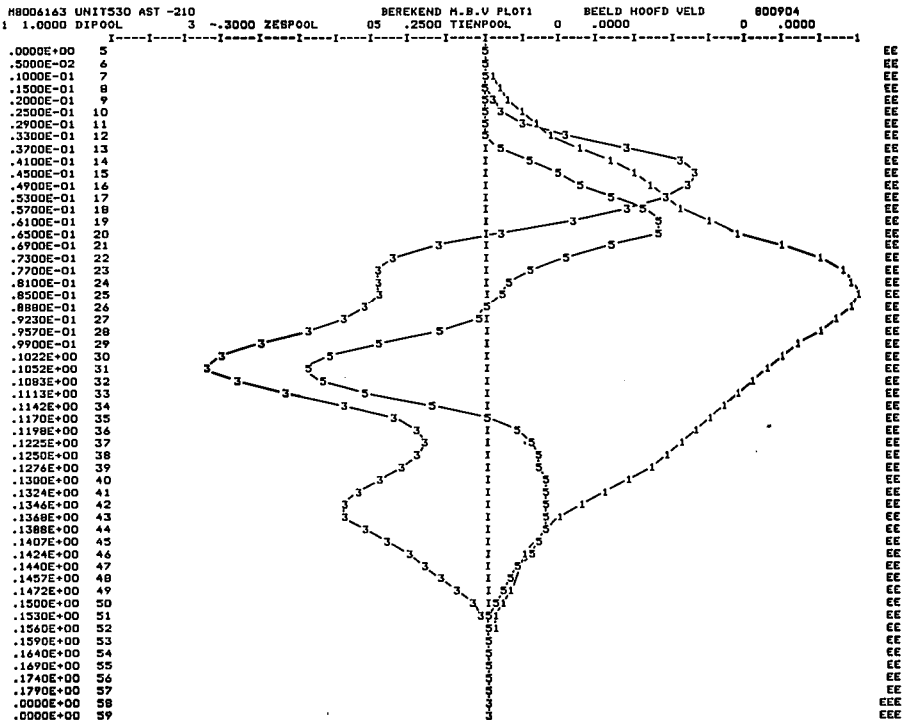


Fig. 28. Example of a multipole composition as supplied by the field-measuring equipment. The location of the numbers 1, 3 and 5 gives the amplitude (horizontal) as a function of position (vertical) along the contour of the tube for the dipole $\mathcal{V}_1$, the sixpole $\mathcal{V}_3$ and the tenpole $\mathcal{V}_5$ of the 30AX frame coil. $\mathcal{V}_3$ is shown on a scale 3.3 times larger than $\mathcal{V}_1$; $\mathcal{V}_5$ is on a scale four times larger.

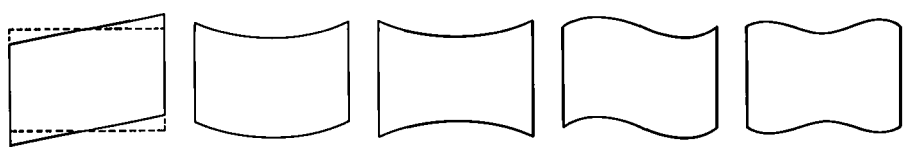


Fig. 29. Examples of simple, easily recognizable deflection errors: NS-distortions of the first to the fifth order. The last one is the familiar 'seagull' distortion.

ascending order. A rapidly varying pattern on the screen suggests a high order (but not every higher-order error gives such a pattern). This recognition process can be made somewhat more systematic by changing one multipole of an experimental design at one position — e.g. at the neck, in the centre or at the screen — and then recording the error produced. This cannot be done with an actual experimental model, but it can be simulated on the computer. We recorded the effects of many multipoles at four positions on the axis in this way. Examples are shown in *fig. 30*.

zation procedure. Such a procedure approaches the optimum multipole composition one step at a time, searching at each step for the particular variation of the multipole composition that reduces the 'error total' as measured by certain predetermined criteria.

We shall now explain this optimization a little more fully. Let the multipole distribution be given by the amplitudes A_k of the multipoles at a number of points on the contour of the tube. Let us take, for example, two-, six-, ten- and 14-poles (two of each) at four places along the contour, making a total of 32 values. Small variations ΔA_k in these amplitudes cause small changes Δa_i in the cal-

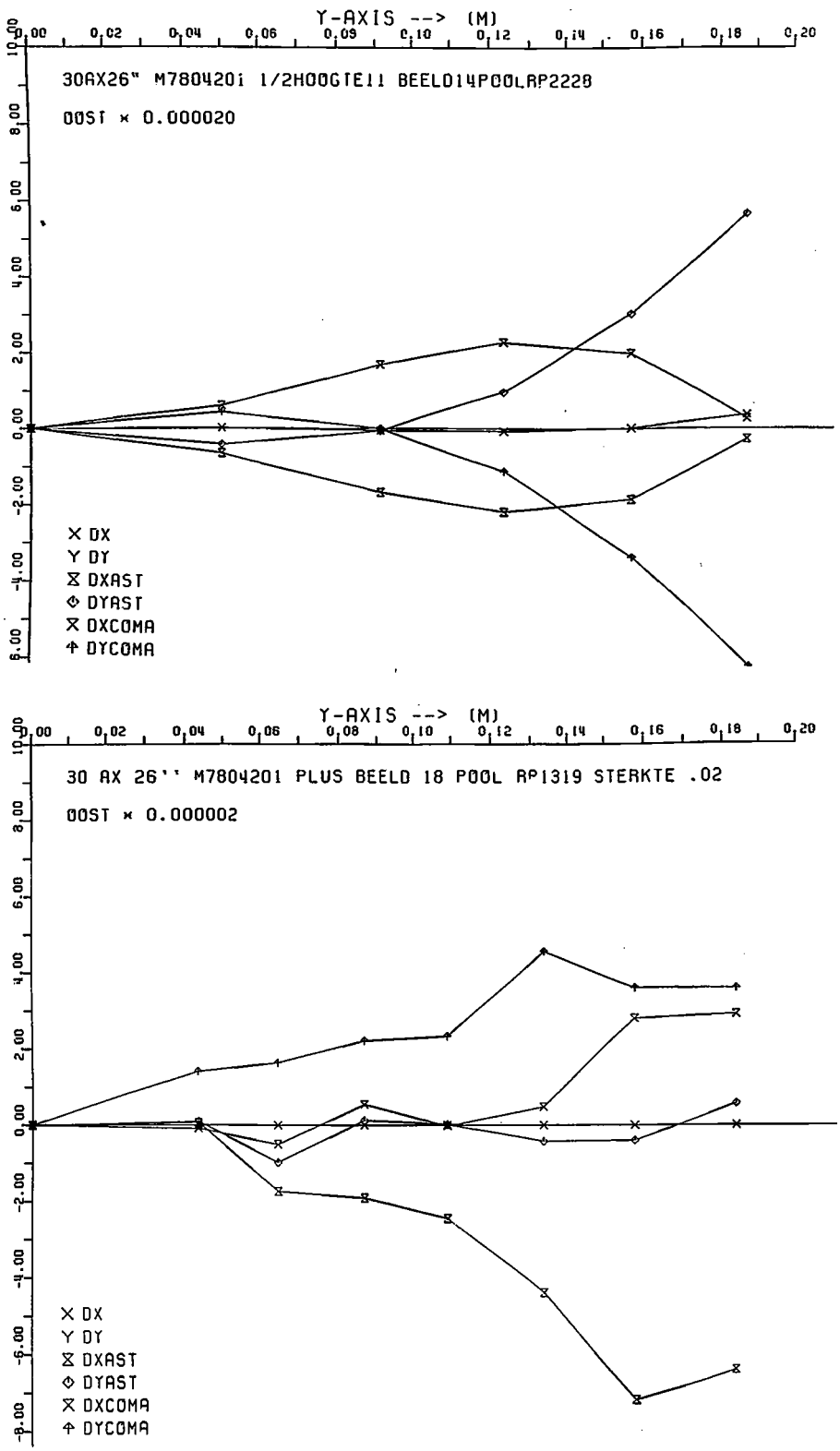


Fig. 30. Deflection errors caused by a multipole at a particular location, as obtained by computer simulation. *Above:* frame 14-pole at the position '22-28'; *below:* frame 18-pole at '13-19'. The errors are on the right-hand side of the screen ('OOST'). The errors are plotted as a function of the coordinate along this edge ('*Y*-axis'). The simulation was carried out for 2-, 6-, 10-, 14- and 18-poles at four positions along the contour ('13-19', '22-28', '31-37' and '40-46').

culated deflection errors. Let us consider the geometrical error, the astigmatism error and coma (six numbers) at say 15 points of a quadrant on the screen, so that there are 90 values. To a first approximation the Δa_i will depend linearly on the ΔA_k:

$$\Delta a_i = \sum_k c_{ik} \Delta A_k.$$

If the errors present are given by the numbers b_i (again 90 in total), we try to compensate for them by means of the changes Δa_i. We thus look for variations such that:

$$\sum c_{ik} \Delta A_k + b_i = 0.$$

It is not possible to find a general solution to these equations, since there are more equations (90 in our example) than unknowns (32). The best we can do is to look for an optimum solution, e.g. by using the method of least squares.

Practical design of the coils

Finding a distribution for the wire and magnetic material that produces the optimum field distribution is usually the biggest problem in the development of a deflection unit. In the first place it is often not possible to work out an 'ideal' or 'optimum' wire and material distribution on paper. A useful rule of thumb has already been mentioned, i.e. that a particular harmonic in the current distribution excites the corresponding multipole (p. 162), so that for a tenpole, for example, the wire density along the circumference should be increased and decreased with a period of 72°. But a wire distribution at a particular position also affects its environment; moreover the wiring must be continuous. Because of such factors a multipole cannot be altered at just one place, nor can it be altered selectively: if we change a tenpole, the sixpole is affected too, which means that we must revise the third-order design again.

In the second place, a wire and material distribution that should theoretically give the desired result may not in practice be feasible. We have already seen one example of this: in the winding technique we use it is not possible to form an indentation in the turns.

Two kinds of situation are therefore encountered in efforts to improve an experimental design. In the first case, the coil parameters available in the particular method used for making the coil can be continuously varied. If the desired field can be produced in this way, the design is ready. If it cannot, then new possibilities and different parameters have to be created, such as the soft-magnetic metal mentioned earlier. Many 'inventions' concerned with deflection coils relate to such ideas.

To conclude, we shall give an example.

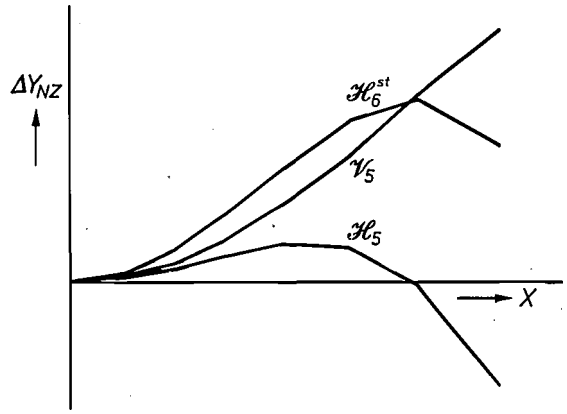


Fig. 31. Effect of $\mathcal{H}_5$, $\mathcal{V}_5$ and a static 12-pole ($\mathcal{H}_6^{st}$) close to the screen on the NS-distortion at the upper edge of the picture, as shown by the computer simulation. As the theory predicts, $\mathcal{H}_6^{st}$ is equivalent to $\mathcal{H}_5$ plus $\mathcal{V}_5$. The $\mathcal{H}_5$ and the $\mathcal{H}_6^{st}$ are used for eliminating the 'seagull' distortion.

'Seagull' distortion

At a particular stage in the development of the 30AX deflection unit there was still an unacceptable NS-distortion of the fifth order: a 'seagull' distortion of about 5 mm (see fig. 29). This is a side-effect of the sixpoles in the third-order design. To keep the design intact, the sixpoles have to be left untouched and the seagull distortion has to be corrected by appropriate means, i.e. by using tenpoles. In our record of multipole effects (see fig. 30) we have *fig. 31*; this shows the effect of $\mathcal{H}_5$ and $\mathcal{V}_5$ close to the screen, at the north edge of the picture. $\mathcal{H}_5$ does in fact give seagull distortion, while $\mathcal{V}_5$ only produces pin-cushion distortion.

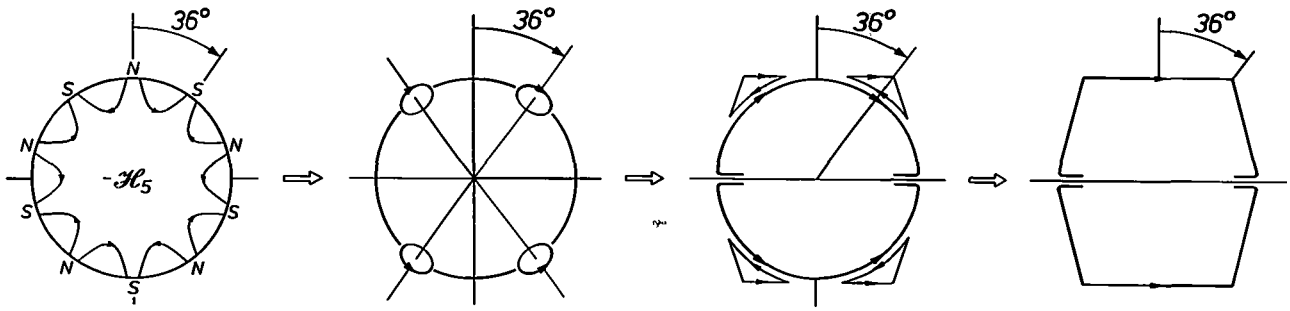


Fig. 32. Diagram showing how a negative $\mathcal{H}_5$ is obtained by means of angular wiring. See fig. 4.

A negative $\mathcal{H}_5$ therefore seems to be the cure for seagull distortion. This treatment must be applied close to the screen, because first the effect on the picture is greatest there, and secondly the unwanted effect on the astigmatism and coma is least there — a situation we have already encountered with the sixpoles.

We produced a negative $\mathcal{H}_5$ by introducing bulges at ± 36° to the vertical in the wiring of the line coil near the screen (see fig. 4); *fig. 32* shows the relation

between wiring and tenpole. The effect on other multipoles is found to be small.

This modification considerably reduces the seagull distortion. We corrected the residual distortion by putting permanent magnets on the screen side, above and below the line coil (see fig. 4). As noted earlier (page 162), a static $2(n+1)$-pole is equivalent to the sum of an $\mathcal{H}_n$ and a $\mathcal{V}_n$. Thus, with a suitable *12-pole* we can enhance the $\mathcal{H}_5$ already made (see fig. 31). By giving the small magnets the appropriate length and strength, we were able to eliminate the seagull distortion in this way.

Even though our corrective measures for the seagull distortion were applied close to the screen, they increase the convergence errors by about 1.5 mm. These are fifth- and higher-order errors additional to the higher-order errors of the dipoles and six-poles. We have used the optimization programs mentioned above to determine the modifications in the multipole composition that would minimize these errors. We were able to make these modifications by dividing the wiring near the screen and neck into separate bundles (see fig. 4). The holes between the bundles are made by 'shooting' pins into the winding gap during the winding process (fig. 25, inset).

Summary. In the 30AX colour-picture system a 'dynamically convergent' unit is produced when deflection unit and picture tube are combined: the tedious setting-up procedures of the past are no longer required. The article deals with the design of the deflection unit. The deflection fields are expanded into multipoles. Dipoles provide the deflection proper ('first order'). Deflection errors of the third order are corrected by means of sixpoles, fifth-order errors by means of tenpoles, and so on. Because of the symmetry there are no even-order errors. The distribution of wiring and soft-magnetic material can be designed in such a way as to provide the rapid variation of the sixpoles along the axis necessary to eliminate all third-order errors. To reduce higher-order errors both the errors and the multipole composition of an experimental model are measured with automated equipment. Comparison of the results with the effects obtained by computer simulation of separate local multipoles shows how to improve the design. Finally, a computer optimization of the multipole composition is used to minimize the remaining unidentified errors. If this cannot be achieved by variation of quantities such as the distribution of wire and soft-magnetic material, new coil parameters have to be introduced. 'Seagull' distortion is given as an example.

Error control in mobile-radio data communication

R. C. French and P. J. Mabey

Mobile radio — an essential for police, fire brigades, ambulances, taxis and public transport — can transmit far more than simple spoken messages. Selective calling, fast and automatic transfer of standard messages, transmission of printed text to and from the vehicle and even of pictures are all possible with the use of digital transmission. However, the mobile-radio channel is plagued with fading and interference, and quantitative information relating to this has been collected in measurement runs in central London. Despite the channel imperfections, reliable data transmission is possible, but it is necessary to use carefully selected codes in which sufficient 'redundant' bits have been added to the information bits to enable errors to be detected and corrected.

Digital signals in mobile radio

Applications

In mobile radio spoken messages are transmitted to and from vehicles. For a long time this was the only function of mobile radio, and it still is in many cases. However, with the increasing use of data processing in binary digital form, it is a natural development to apply the same methods to mobile radio. The possibilities are many and varied.

One example is a selective calling system. When a vehicle is called, the base station first transmits the digital code for that vehicle. The receiver in the vehicle reacts by switching on the audio circuits, so that the call can be heard. The other receivers remain quiescent and no-one is distracted by messages intended for others. In addition, the mobile station always gives its own identity code automatically at the start of every transmission; this ensures that the transmitting vehicle is identified at the base station. Similar systems, but working with combinations of audio tones, are on the market [1].

Digital signalling can also be used to report the status of the vehicle. The mobile operator can do this by setting a selector switch on his equipment to the appropriate position — e.g. 'busy', 'free', 'on watch', 'off watch' — and the base station periodically requests the status of all vehicles, which then reply automatically. Such a system is also on the market [2]. In addition, this system can send a hundred precoded and numbered messages; the operator chooses the appropriate number and presses the *SEND* button (*fig. 1*).

Fig. 1. Pye MDU 1000 mobile data unit, for operation with the DS 1000 vehicle-availability system. A hundred pre-numbered messages can be sent (*SEND*) and received (*MESSAGE*). The status of the vehicle can be reported by using ten other pushbuttons — 'busy', 'free', 'on watch', etc.; these messages are sent automatically in response to a call from the base station. An identity-code card is inserted in the slot at lower left.

R. C. French, Ph.D., and P. J. Mabey, B.Sc. (Eng.), are with Philips Research Laboratories (PRL), Redhill, Surrey, England.

Data transmission is also used for vehicle location by radio [3]. Moreover, it can increase the information flow to the vehicle. The vehicle can be provided with a printer unit that will reproduce maps, pictures, etc.; matrix printers are very suitable here (*fig. 2*). A computer terminal can be fitted in the vehicle for direct interrogation of a data base; police in Canada are already using such methods in detective work. Finally, we should mention here the application of digital transmission in larger systems for calling individual persons by radio (e.g. digital radio paging schemes in London and Chicago) [4] and for dialling telephone connections between a vehicle and the public telephone system.

Reception conditions and interference

In the United Kingdom, the mobile-radio frequencies are about 173 MHz in the VHF band and about 462 MHz in the UHF band, i.e. wavelengths of about 1.7 m and 65 cm. Narrow-band frequency modulation is used (the channel spacing is 25 kHz at UHF and 12.5 kHz in the VHF band).

One of the problems in mobile radio is the economic use of the frequency spectrum. In the Philips research organization and elsewhere efforts are being made to find digital modulation systems that give a minimum bandwidth [5]. The methods so far used in digital mobile-radio transmission include frequency-shift keying (FSK) of a subcarrier [6].

Another problem is the corruption of the digital signal by noise of various origins, and we shall consider this here. Mobile radio has to operate under extremely unfavourable conditions. The transmitter powers are low, e.g. 20 W from a base station (5 W from a mobile unit), and the transmitting antennas are not very high, e.g. 30 m above the ground. This means that the field-strength is small and the mean signal voltage at the receiver antenna input may often be only a few microvolts. Multiple reflections of the signals by hills and buildings set up a standing-wave pattern, so that there are rapid variations in signal strength ('fading') as the vehicle moves through the pattern; the signal level often changes by more than 20 dB in less than a metre. In addition to the reflections there is the shadowing effect of hills and tall buildings, which lowers the mean signal level locally.

Besides poor propagation the received signal is contaminated by ignition noise from nearby vehicles and with 'co-channel interference' resulting from re-use of the same radio channel in another part of the country. The ignition noise results in a significant number of errors in the received data signal, particularly at higher bit rates in the lower-frequency mobile-radio bands. Co-channel interference resulting from channel re-use is significant because of the great variability in both the wanted and the unwanted signals caused by multipath propagation, which makes it necessary to use large re-use distances [7]. However, a shortage of radio channels encourages the use of the minimum re-use distance possible.

Fig. 2. As shown by this picture of Marilyn Monroe, a generally useful identity portrait can be obtained by printing out the image, transmitted in digital form by the mobile-radio system, on a matrix printer. Maps, etc., can also be rapidly made available to the vehicle crew in the same way.

[1] An example is the UED 6 universal encoder-decoder unit made by Pye Telecommunications Ltd, Cambridge, England.

[2] Pye DS 1000 vehicle availability system.

[3] R. W. Gibson, PINPOINT — a radio system for locating and monitoring vehicles, Philips tech. Rev. **35**, 15-22, 1975.

[4] R. H. Tridgell, The application of coding techniques to radio-paging, in: 1980 Int. Zurich Seminar on Digital Comm., pp. C 9.1-9.5.

[5] L. E. Zegers and C. B. Dekker, A comparison of digital transmission techniques for standard FM mobile radio sets, IEEE Trans. **COM-25**, 364-368, 1977.
F. de Jager and C. B. Dekker, Tamed frequency modulation, a novel method to achieve spectrum economy in digital transmission, IEEE Trans. **COM-26**, 534-542, 1978.

[6] H. C. van den Elzen and P. van der Wurf, A simple method of calculating the characteristics of FSK signals with modulation index 0.5, IEEE Trans. **COM-20**, 139-147, 1972.
An explanation of FSK is given in F. W. de Vrijer, Modulation, Philips tech. Rev. **36**, 305-362, 1976.

[7] R. C. French, The effect of fading and shadowing on channel reuse in mobile radio, IEEE Trans. **VT-28**, 171-181, 1979.

These poor reception conditions cause a high bit-error rate (e.g. 1%) with digital signals. To permit an accurate estimate of the transmission quality to be obtained a good statistical description of the errors is required. We therefore made measurements in a vehicle, both in London and near the laboratories at Redhill [8]. Tape recordings were made of the error patterns in real conditions to investigate various error-detecting and error-correcting codes. It was found that when a code with sufficient added 'redundant' bits was used, satisfactory data transmission was possible even on the difficult mobile-radio channel. We shall look at this in more detail later. Before that, however, we shall give an account of the measurements in a moving vehicle and we shall discuss the error rates at various signal levels and various bit rates.

Measurements in London

The measurements in London enabled us to collect statistical data about the field-strength distribution and the bit errors introduced in digital transmission. The measurements were made at 17 locations (*fig. 3*). The transmitter was located on the roof of Mullard House, a ten-storey office building in central London, and had a power of 3.5 watts; the transmissions were made in the UHF band at a frequency of 462.425 MHz. The continuously transmitted data signal was a pseudo-random sequence consisting of a repeated cycle of 127 bits.

The test vehicle carried equipment for recording the signal level at the antenna and the errors in the received data signal. The signal level was sampled periodically during the test runs and the dB values obtained with the aid of a logarithmic amplifier were recorded on a cassette tape. This data was used later for calculating mean values and statistical distributions on a minicomputer. *Fig. 4* shows a number of mean values, each relating to a test run of 7 minutes at a certain distance from the transmitter. For comparison we have included a curve showing the signal level to be expected for propagation in free space. The measured values were 30 to 50 dB below this curve! Similar results have been found elsewhere [9].

The errors in the received data signal were detected by synchronously generating the same 127-bit pseudo-random sequence locally in the vehicle and comparing the local and received sequences. The outcome of this comparison was recorded bit by bit on a digital cassette tape. Besides single isolated bit errors, mainly due to ignition noise, there are short periods (typically ten milliseconds) during which many of the bits are in error. These are known as bursts of errors and are the result of passing through a fade. If it is assumed that the fades are half a wavelength apart and that the vehicle travels at 30 km/h in town traffic, this means that it passes through about 26 fades per second.

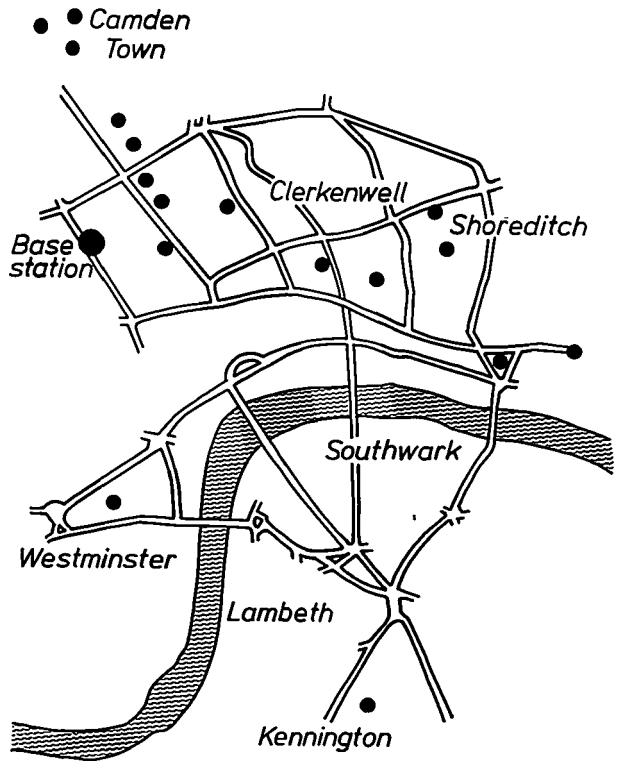


Fig. 3. The map of central London, showing the 17 locations (•) where the field-strength and bit-error measurements were made in a moving vehicle for mobile-radio studies. The transmitter (*Base station*) was located on the roof of Mullard House, a ten-storey office building near Tottenham Court Road.

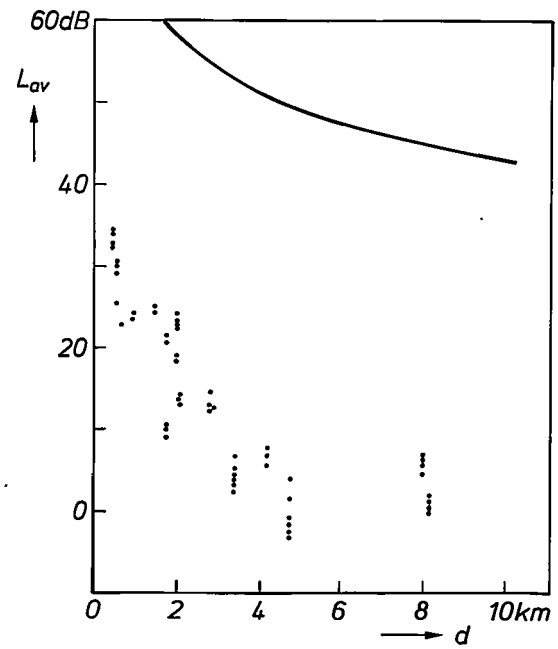


Fig. 4. *Measured points:* level L_{av} of the received signal, averaged over a run of seven minutes at a distance d from the transmitter. 0 dB corresponds to 1 μV at the antenna terminals. *Curve:* the signal level as it would be for propagation in free space.

The number of errors due to noise and fading would be expected to depend on the bit rate. Measurements were therefore made at various bit rates: 37.5 bits/s, 1200 bits/s and 4800 bits/s. The low bit rate of 37.5 bits/s was included in the experiments because it is used in selective calling and paging systems. However, if more data has to be transmitted this bit rate is too low; the Pye DS 1000 system mentioned earlier operates with a bit rate of 1200 bits/s.

The number of bit errors is found to increase markedly when the duration of a single bit becomes shorter than the time taken to go through a fade. This was demonstrated most clearly in experiments made in the immediate neighbourhood of our laboratories at a field-strength high enough for the effects of ignition noise to be negligible. In these suburban surroundings the vehicle can travel faster — at about 45 km/h (28 mph) — i.e. about 40 fades per second. It can be seen from *fig. 5* that the bit-error rate becomes about three orders larger when the bit rate increases from 37.5 to 150 bits/s, with very little further increase above 150 bits/s.

The same trend was found in the measurements made in London. A comparison of the three bit rates mentioned above can be made from *fig. 6* [10], but here the error rate has been plotted against the mean signal level. Besides fading and shadowing, ignition noise also played a part here. The curves shown are based on means for a seven-minute run. The curves for 1200 bits/s and 4800 bits/s have a slope that corresponds approximately to a decrease in the bit-error rate by a factor of ten for a 10 dB increase in the mean signal level (i.e. a tenfold increase in power). The curve for 37.5 bits/s is much steeper. Even at a signal level of less than 10 dB (with respect to 1 μV) the bit-error rate has fallen to less than 10^{-5}. This means that systems that operate with this low bit rate, such as those for selective calling, simple position reporting and other short messages, can be sufficiently reliable without the use of error-control codes.

This is not the case at higher bit rates. To reduce the error rate to an acceptable value (10^{-5} to 10^{-6}) error-detecting or error-correcting codes will have to be used here. For these codes it is necessary to group the bits in words or 'blocks'. It is then important to know how many errors might be expected in a single block, since this number is the critical factor in the choice of a code. For blocks of 64 bits this number can be read

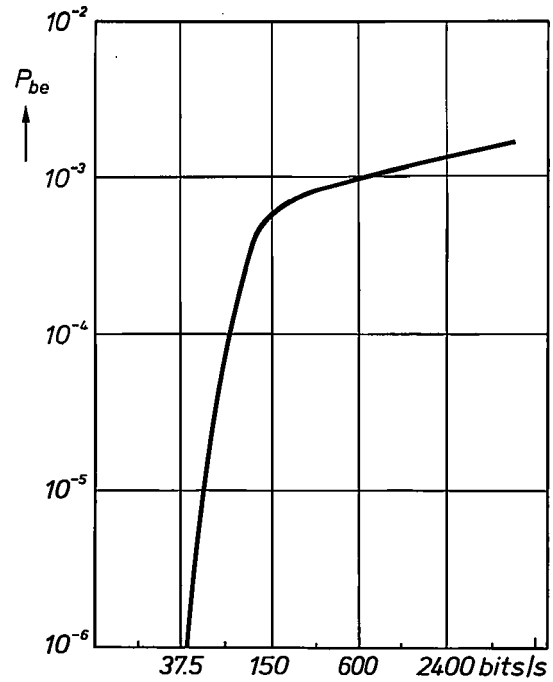


Fig. 5. Bit-error rate P_{be} measured as a function of bit rate. P_{be} increases rapidly as the bit period becomes shorter than the duration of a fade. The vehicle speed was 45 km/h; there were about 40 fades per second. Signal level 10 dB (with respect to 1 μV), no ignition noise.

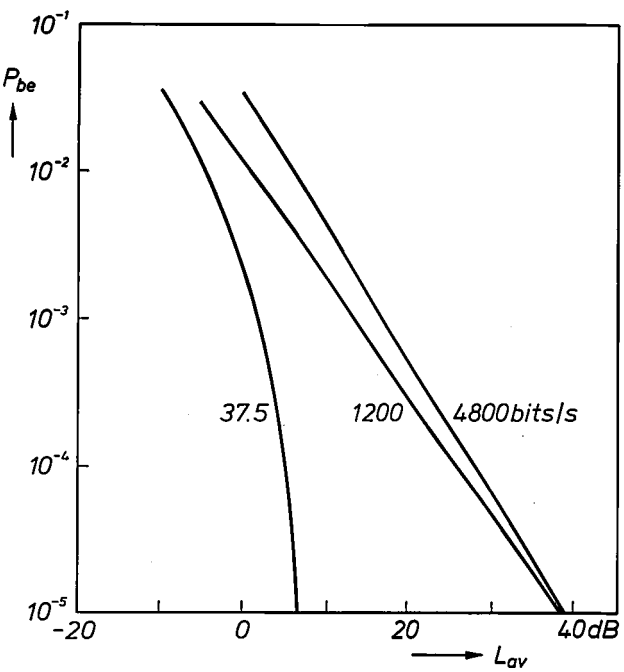


Fig. 6. The bit-error rate P_{be} as a function of the mean signal level L_{av}, measured in London. At higher bit rates the error rate decreases by a factor of about 10 for an increase in level of 10 dB. At 37.5 bits/s the duration of a single bit is longer than the time taken to pass through a fade; as soon as the signal is strong enough to override the ignition noise a very reliable transmission is obtained.

[8] R. C. French, Radio propagation in London at 462 MHz, Radio and electronic Engr. 46, 333-336, 1976.
R. C. French, Mobile radio data transmission in the urban environment, in: IEEE 1976 Int. Conf. on Communications (ICC 76), Philadelphia, pp. 27.15-27.20.

[9] E.g. in Tokyo: Y. Okumura, E. Ohmori, T. Kawano and K. Fukuda, Field strength and its variability in VHF and UHF land-mobile radio service, Rev. Electr. Comm. Lab. 16, 825-873, 1968.

[10] R. C. French, Error performance in mobile radio data transmission in the urban environment, NTG-Fachber. 61, 83-88, 1977.

off from the cumulative error distribution in *fig. 7*, which is calculated from the measured results at a bit rate of 4800 bits/s. It can be seen from this figure, for example, that at a mean signal level of 11 dB (with respect to 1 μV) 1% of the blocks have eight or more errors. At 0 dB 1% of the blocks have 21 or more errors and one block in a thousand has as many as 32 errors, the same number that would occur if no signals at all were received.

Principles of coding for error control

Error-control coding is used to prevent false messages from being accepted, by detecting errors, or to enable messages to be accepted despite the occurrence of bit errors, by correcting errors. When an error-detecting code is used, code words containing errors are discarded. This results in some characters being left out when the message is printed. The code words may be repeated to complete the message. Error-correcting codes, on the other hand, will correct some of the errors so that fewer repeats are necessary.

Without error-control coding, every combination of received bits is a valid message, and bit errors will transform a transmitted message into a different, false message. With an error-control code, the bit sequences transmitted are restricted so that only certain combinations of bits form valid messages. This is similar to written language where not all combinations of letters form valid words. The bit patterns of the code, the code words, are chosen so that when bit errors occur, the resulting bit sequences rarely form valid code words. Error detection consists in recognizing when invalid words have been received. This is similar to detecting errors in written language by determining whether the words are in the dictionary. Error correction consists in estimating which code word was transmitted when an invalid code word is received, which is like locating the most similar word in the dictionary. Comparison of an invalid word with the most similar valid word enables the bit errors to be located and corrected.

In order to restrict the bit patterns which are transmitted without restricting the messages which may be sent, redundant bits must be transmitted. Check bits are appended to information bits to form code words. Coding theory provides rules for calculating the check bits from the information bits, ensuring that the resulting code words are dissimilar but without adding an excessive number of check bits. The more dissimilar code words are, the less likely it is that sufficient errors will occur to transform one code word into another code word, so the less likely false messages will be.

For example, if code words differ in at least three bit positions, then at least three bit errors are necessary to cause a false message which will pass undetected. Furthermore, if only one bit error occurs, then the corrupted code word will be more similar to the transmitted code word than to any other code word, and so the error can be corrected. If a decoder corrects errors, then, in our example, fewer than three errors are sufficient to cause a false message. If two bit errors occur, then the corrupted code word may be more similar to a false code word than to the transmitted code word.

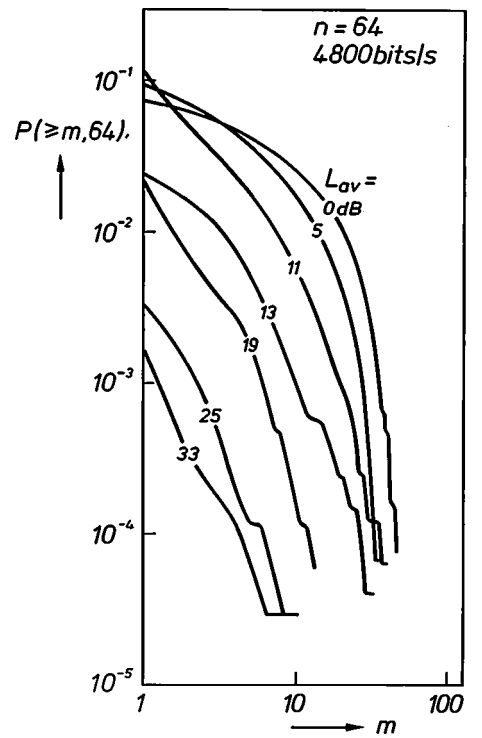


Fig. 7. The probability $P(\geq m,64)$, that the number of errors in a code word of 64 bits is equal to m or more, for various mean signal levels L_{av}. The probability was calculated from the measurements made in London and is of interest for the application of error-detecting or error-correcting codes.

In coding jargon, the number of bits by which the most similar code words in a code differ is called the 'minimum distance' of the code and is normally denoted by d. When fewer than d errors occur we can be sure of detecting the errors, when d or more errors occur there is a risk that a false message is formed. If we use an error-correcting decoder we can be sure of successfully correcting $(d-1)/2$ or fewer errors. With more errors there is a risk of a false message. The chance of a false message can be calculated [11].

As well as providing rules for calculating the check bits from the information bits, coding theory also provides simple algorithms for implementing en-

coding. A simple encoder also means a simple error-detecting decoder. As mentioned above, error detection consists in determining whether the received information and check bits form a valid code word, i.e. whether the received check bits are the same as are obtained by re-encoding the received information bits. If there are errors in either the information or the check bits, re-encoding results in a different set of check bits from those received. It is not possible with an error-detecting decoder to know whether the errors are in the information or the check bits.

Error correction is more complex to implement than error detection, but again coding theory provides algorithms. In 'cyclic block codes', which are the ones considered in this article, each check bit is an overall parity check on a different set of the information bits such that each information bit is cross-checked. Each check bit is chosen so that the information bits and the check bit together have an even number of binary ones. By noting which of the parity checks fail, i.e. by noting which of the recalculated check bits differ from the received check bits, it is possible to determine which bits of the code word are in error and correct them [12].

In the following we shall represent the number of bits in a code word by n, and the number of information bits in it by k, so that there are $n-k$ check bits in each code word. An error-detecting code will be designated as an (n,k) code for convenience; in a (15,7)-code for example each code word consists of 15 bits and includes 7 information bits. An error-correcting code will be designated as an (n,k,t) code, where t indicates the number of bit errors that can be corrected per code word.

The simplest example of an error-detecting code is a code in which each word of n bits consists of $n-1$ information bits and one check bit. The value of the check bit is chosen to ensure that the word contains an even number of binary ones (even parity). Alternatively odd parity could be used. Now if a single error occurs in the transmission (in an information bit or in the check bit), the parity of the received word will not correspond to the prearranged parity, showing that there is an error. It is clear that this $(n,n-1)$ code is only able to detect odd numbers of errors in a code word; it is therefore of no use in mobile radio. An error-detecting code for use in mobile radio will have to contain more check bits.

If it is desired not only to detect an error but to correct it as well, then it is necessary to find out first if there is an error and then which bit is incorrect. With a single error in a code word of n bits this amounts to $n+1$ unknowns. The $n-k$ check bits must be able to supply at least this amount of information, so that for the correction of a single error the condition necessary is

$$2^{n-k} \geq n+1.$$

The number of check bits decides the number of check operations that have to be made. The check mechanism consists of a set of equations in which suitably chosen bits from the code word are summed modulo 2 to check for even parity (or odd parity if desired). Each check operation provides one bit; these bits form a number called the syndrome. If there are $n-k$ check bits, the syndrome also consists of $n-k$ bits. The value of the syndrome can be used to find the position of the error.

Choosing the optimum code

The choice of a code for any data-transmission system should take into account the characteristics of the communications channel. In some cases, the performance of a code may be calculated under the assumption that no clustering of errors occurs, but that all errors are independent. This approach permits a straightforward mathematical prediction of performance but can give very misleading results for mobile radio. For example, a code which, according to independent error calculations, enables 99.999% of code words to be received correctly, in fact results in only 99% of the code words being correct when the code is used for mobile radio. It is therefore important to know what the error distribution is, and this information is contained in the error-pattern recordings.

One approach used for assessing the performance of a code is to build an encoder and decoder in hardware or software and to transmit data between them in the laboratory, permitting the real error patterns to corrupt the data before it is decoded. This method has its limitations, as it leads to lengthy measurements, particularly if very small probabilities are to be measured. If, for instance, only one false message in a million occurs in a coding scheme then many millions of messages will have to be processed in order to measure the false rate. This requires very long error-pattern recordings, far longer than could be collected conveniently while driving round a city.

We adopted a method of code evaluation which overcomes the problems described above by analysing the error pattern statistically. Performance can then

[11] R. W. Lucky, J. Salz and E. J. Weldon, Jr., Principles of data communication, McGraw-Hill, New York 1968, pp. 355-360.
P. J. Mabey, Mobile radio data transmission — coding for error control, IEEE Trans. VT-27, 99-109, 1978.

[12] Coding theory and error-correcting algorithms have been studied extensively at the Philips research laboratories in Brussels; see for example:
J.-M. Goethals, Combinatorial decoding methods for block codes, in: Théorie de l'information, Coll. Int. CNRS No. 276, Paris 1977, pp. 223-231;
J.-M. Goethals, Single-channel error-correcting convolutional codes, Philips J. Res. 33, 248-253, 1978.

[13] Another approach would be to model the error distribution mathematically and use the mathematical model to analyse the performance of codes. This approach was used at the Philips research laboratories in Brussels to evaluate codes for correcting bursts of errors occurring on computer magnetic discs. Results can be found in: Ph. Piret, Structure and error probability of burst-correcting convolutional codes, Philips Res. Repts. 27, 244-256, 1972.

be simply calculated, different codes can be quickly assessed, and the effect of changing the code parameters seen [13].

Error-pattern analysis and code-performance calculation

To assess the performance of a code we need to know how often a code word is corrupted by errors and how often the error-detecting or error-correcting ability of the code is exceeded. To find this out, we divided the measured error patterns into blocks of bits and counted the number of errors in each block. The results of this are plotted as the probability $P(\geq m,n)$ of finding m errors or more in a block of n bits, as a function of m and with n as a parameter. Fig. 11 (discussed more later) illustrates the $P(\geq m,n)$ distributions (solid lines) for a range of block lengths n (from 7 to 255 bits) based on error patterns we measured at a mean received signal level of 1.9 dB (with respect to 1 μV). It shows, for example, that at this signal level 2% of 15-bit blocks will contain three or more errors.

We used the $P(\geq m,n)$ distributions to calculate the performance of codes in terms of their success rate P_c and false rate P_e. Only cyclic block codes were considered. These codes have also been widely investigated elsewhere, and many powerful codes with simple encoding and decoding procedures are known [14].

For error-*detecting* codes, the following equations are used in the calculations:

$$P_c = 1 - P(\geq 1,n), \quad (1)$$

$$P_e \approx 2^{-r} P(\geq d,n). \quad (2)$$

Here P_c is simply one minus the probability of any errors $P(\geq 1,n)$. In eq. (2), r is the number of redundant bits in each code word. Remembering that all code words differ from each other by at least d bits, we can derive the approximate equation for P_e as follows. An error-detection decoder tests for errors by checking whether a received block is a code word, i.e. if the r check bits are those that follow from the k information bits. Now the probability that the r check bits are right by chance is 2^{-r}, and this is the first factor in (2). Secondly, the error pattern must contain at least d errors to corrupt one code word to another, hence the $P(\geq d,n)$ factor.

For error-*correcting* codes, the following equations apply. A decoder is assumed that corrects all patterns of t or fewer errors, and detects all other correctable patterns.

$$P_c = 1 - P(\geq t+1,n), \quad (3)$$

$$P_e \cong P(\geq d-t,n)\, 2^{-r} \sum_{j=0}^{t} C_j^n. \quad (4)$$

Here P_c is one minus the probability of $t+1$ or more errors. Again, the equation for P_e comprises two factors as follows. If all the 2^n possible received bit patterns were corrected to a valid code word, then for each code word there would be $2^n/2^k = 2^r$ correctable error patterns. However, only patterns of $j \leq t$ errors are corrected. For every integer j there are C_j^n possible patterns of errors [15]; the fraction of the 2^r correctable error patterns which the decoder will actually attempt to correct is therefore

$$\sum_{j=0}^{t} C_j^n / 2^r.$$

Further, for an erroneous decoding $d-t$ or more errors must occur.

Throughput

The value of a code cannot be assessed in terms of its success and false rates alone. Throughput, i.e. the number of information bits transmitted per second, is a very important parameter from a practical point of view, and will become even more so as the mobile-radio spectrum becomes more crowded and more channel sharing is introduced. Throughput will suffer if more check bits are added to increase the success rate or lower the false rate or both.

Error-detecting codes

Error-detecting codes are of interest because very powerful codes are available which are easily implemented, in both hardware and software. Integrated circuits are available which perform encoding and decoding of commonly used error-detecting codes.

As an example, the encoding of ASCII characters is considered. The ASCII code is a seven-bit code for alphanumeric characters; it is identical with the international ISO 7-bit code [16] and it can be used for communicating with a mobile printer or visual-display unit and for sending messages from a mobile keyboard.

A single seven-bit ASCII character could be encoded in a (15,7) code. *Fig. 8* shows the success rate and false rate of this code, calculated using the real error patterns, for a range of mean received signal levels. At a typical mean signal level of 10 dB about 99% of the code words are received error-free so about 99% of the characters would be correctly printed or displayed. Most of the other code words have the errors in them detected so these characters cannot be printed or displayed. A space could be left or a special null character substituted. Only one in a million code words at 10 dB have errors which pass undetected and will result in an incorrect character. As the signal level is increased the success rate improves only gradually, an increase of 14 dB being required to improve the success rate from 99% to 99.9%. An increase in level of 14 dB is equivalent to a signal-power increase by a factor of 25. So increasing the transmitter power is not an efficient way of improving the performance.

Better coding provides a more useful solution. The repetition of code words, the use of a higher bit rate and the use of an error-correcting code are three ways of getting a better success rate.

Repetition for improved success

To ensure a very high success rate code words can be transmitted more than once. A simple scheme would be to send every code word twice. On a burst-error channel, if a code word is corrupted by errors then a repeat sent immediately after may be corrupted by the same burst, so for maximum benefit the repeat should be delayed. A delay of a few code-word lengths at 1200 bits/s is typically sufficient to make this burst effect negligible [17].

Automatic repetition of every code word will halve the throughput on the channel. If a return channel is available the error-detection decoder can request retransmission of just those code words that were corrupted by errors so that the throughput is negligibly degraded under normal conditions.

Higher bit rate

Data-transmission equipment currently being introduced for mobile radio commonly uses a data rate of 1200 bits/s. However, the mobile-radio channels, which, in the UK, are spaced at 12.5 kHz in the VHF bands and at 25 kHz in the UHF bands are capable of supporting data rates of 4800 bits/s or 9600 bits/s or more respectively, using direct frequency modulation of the radio-frequency carrier, without the introduction of a subcarrier, which is customary at 1200 bits/s. These higher bit rates result in longer bursts of errors than at 1200 bits/s, so an error-detecting code will have a poorer performance when used at these higher bit rates. But it is important to know whether, by using more redundancy, a system operating at a higher bit rate can equal the performance of a 1200 bits/s system and in addition achieve a higher throughput.

A comparison was made using the UHF measurements at 4800 bits/s and 1200 bits/s. The performance of a (15,7) error-detecting code is plotted in *fig. 9* for both bit rates. The success rates in each case are not very different, while the false rates are, at 10 dB, almost an order of magnitude worse at the higher bit rate. This difference in false rate may be made up for by using three or four additional bits of redundancy at 4800 bits/s. For example the false rate of a (19,7) code is also shown in fig. 9*b* and is better than the (15,7) code at 1200 bits/s. The throughput is 560 information bits per second by sending the (15,7) code at 1200 bits/s, and more than three times this by sending the (19,7) code at 4800 bits/s. Transmitting data at 4800 bits/s therefore allows a significantly

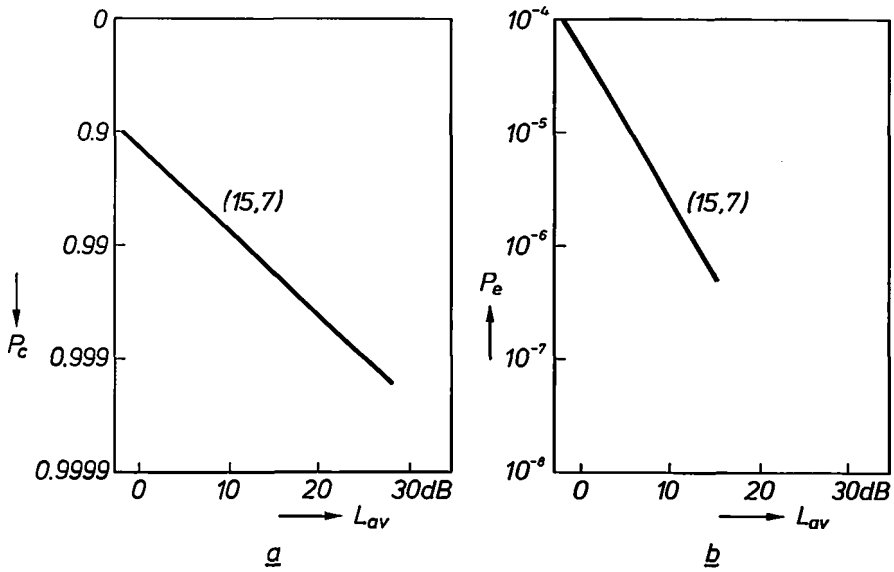


Fig. 8. Success rate P_c (*a*) and false rate P_e (*b*) when a (15,7) code for error detection is used (each code word consists of 7 information and 8 check bits). The success and false rates have been plotted against the mean received signal level L_{av} (0 dB ≙ 1 μV).

[14] W. W. Peterson and E. J. Weldon, Jr., Error-correcting codes, MIT Press, Cambridge Mass. 1972.

[15] C_j^n is the binomial coefficient, also denoted by $\binom{n}{j}$.

[16] ASCII: *A*merican National *S*tandard *C*ode for *I*nformation *I*nterchange. The code conforms to the international standard ISO 646 (1973).

[17] R. C. French, The mobile radio data channel, in: 1980 Int. Zurich Seminar on Digital Comm., pp. D1.1-1.9.

higher throughput to be obtained for the same reliability as at 1200 bits/s, and the error-detection decoder need not be much more complex.

Error-correcting codes

Error-correcting codes provide a means of obtaining a high success rate in systems which are restricted to a low bit rate where error-detecting codes with repeated code words would not give sufficient throughput.

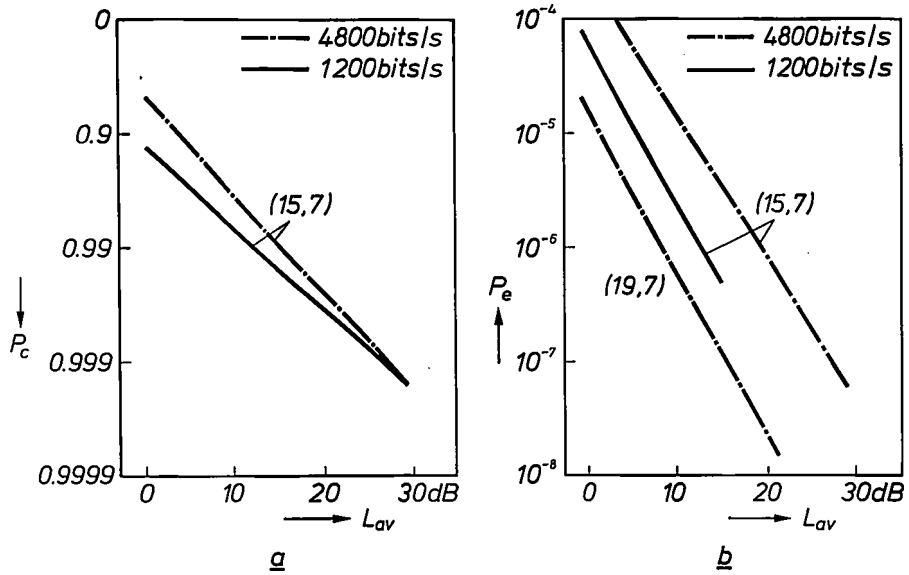


Fig. 9. Success rate P_c (*a*) and false rate P_e (*b*) when a (15,7) error-correcting code is used at two bit rates. The higher bit rate increases the throughput. The deterioration in P_e at the higher bit rate can be counteracted by adding four extra check bits to each code word. Even with the extra check bits the throughput is still three times greater at the higher bit rate.

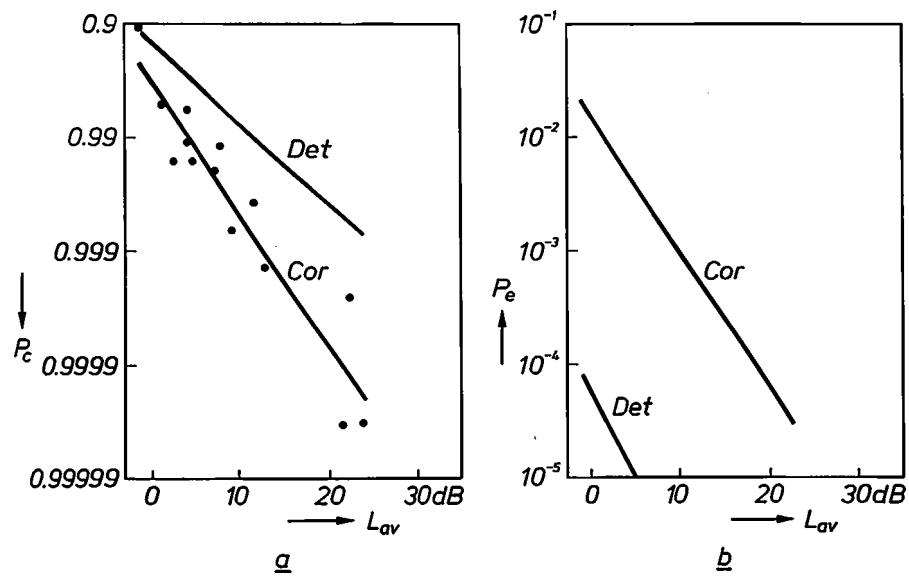


Fig. 10. Comparison of an error-correcting code (*Cor*) with an error-detecting code (*Det*). Both cases refer to a (15,7) code; this can be used for correcting two errors per code word. The success rate P_c is increased by the introduction of error correction (*a*), but the false rate P_e also increases — by a factor of more than 100 (*b*). The measured points in (*a*) refer to the error-correcting code; they give an impression of the typical spread in this kind of measurement.

One obstacle which will delay the introduction of high-bit-rate mobile data systems is the requirement for the data signal to pass from the base station through a telephone line to the transmitter site. High-bit-rate modems for use on telephone lines are expensive because multilevel signals are necessary and amplitude/delay equalizers are usually required. For this reason mobile-data schemes will be limited to 1200 bits/s until costs drop due to LSI.

When discussing error detection the (15,7) code was used to illustrate the performance of an encoded ASCII character. The same code could alternatively

be used for correcting up to two errors. Its success and false rates are plotted in *fig. 10* together with the curves repeated for error detection. The success rate at 10 dB is improved from 99% to about 99.8% by using error correction instead of detection, but the false rate has been degraded by more than two orders of magnitude. For some applications the error correction might be considered superior; for example when transmitting text to a mobile printer the few wrong characters printed when the error-correcting decoder produces a false character may be more acceptable than frequent clusters of characters not printed because of detected errors.

Error-correcting codes therefore enable a better success rate to be obtained at the expense of a high false rate. The other penalty which may be incurred is a complex decoder. Single-error correcting codes can be implemented easily, but codes for correcting more errors rapidly become more complex. On the other hand, powerful microcomputers are now available which make the implementation of quite complex error-correction algorithms feasible. Because the complexity of implementation is so sensitive to the parameters of an error-correcting code it is important to know just how much benefit is gained in performance from using long codes capable of correcting many errors.

The effect of choosing different code parameters can be seen in the $P(\geq m,n)$ distributions. *Fig. 11* shows the $P(\geq m,n)$ distributions for one representative measurement. The points marked on the curves refer to a selection of codes. Codes have been selected having code rates R of roughly $\frac{3}{4}$, $\frac{1}{2}$ and $\frac{1}{5}$, which means that the number of information bits is the given proportion of the total number of bits in a word. To illustrate this consider the (15,7,2) two-error correcting code, whose success rate is $1 - P(\geq 3,15)$; in fig. 11 the $P(\geq 3,15)$ point has been marked on the $n = 15$ curve, and the value of $1 - P(\geq 3,15)$ may be read from the right-hand axis, i.e. $P_c = 0.98$.

The points for the codes at each code rate have been joined by a dashed line, e.g. for $R \approx \frac{1}{2}$ the dashed line joins the points for the (7,4,1), (15,7,2), (31,16,3), ... codes. These dashed lines show that increasing the code-word length (with R constant) produces little change in P_c. Changing the code rate (with n constant) causes only a gradual change in P_c. Some improvement can be gained by using longer words at a low code rate, e.g. $\frac{1}{5}$, but using longer words at a high code rate, e.g. $\frac{3}{4}$, can degrade P_c.

The false rate (not illustrated) is more sensitive to changes in the word length or code rate. A long code with a low code rate is necessary to get a combined high success rate and low false rate.

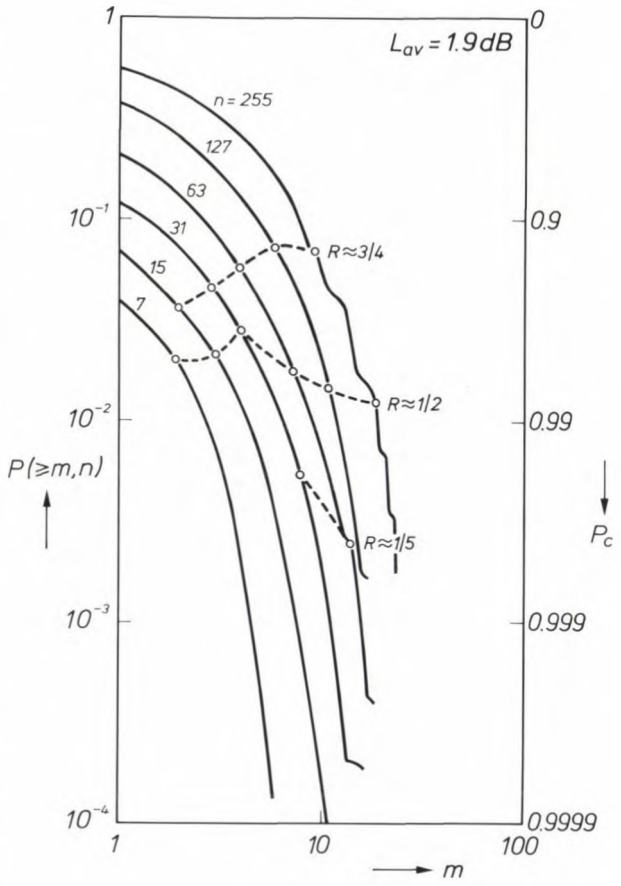


Fig. 11. The probability $P(\geq m,n)$ of m or more errors in a code word of n bits. The curves are based on measurements made at a mean signal level of L_{av} = 1.9 dB. If a code can correct up to $m - 1$ errors, the success rate P_c is $1 - P(\geq m,n)$ (*right-hand ordinate*). A number of such codes with the same code rate R (ratio of the number of information bits to the total number of bits) have been joined by dashed lines; increasing the word length is found to have no great effect on the success rate.

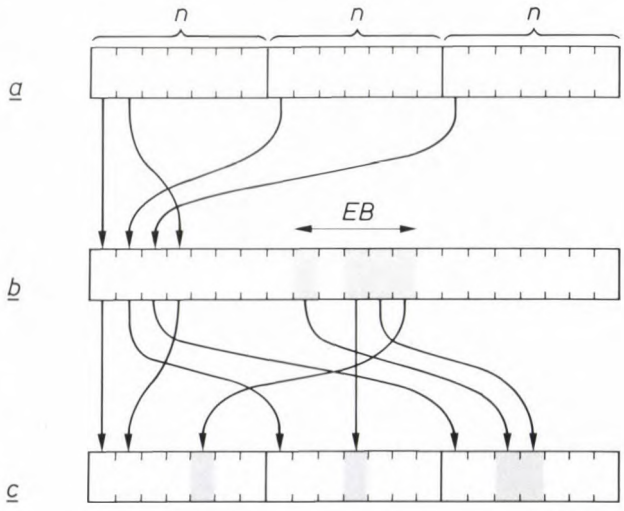


Fig. 12. Three code words of n bits (*a*) are interleaved before transmission (*b*). The bit errors resulting from an error burst *EB* are distributed over three code words in the reconstruction (*c*), so that the number of errors does not exceed the correcting capability of the code for any of the three words. In practice such an interleaving pattern usually contains several hundred bits.

Bit interleaving

A very effective way of coping with clustered errors is to use bit interleaving to disperse the errors so that the errors are less likely to exceed the correcting capability of the code. Before a message is transmitted the bits from several code words are interleaved. Then when an error burst occurs, the errors will be shared among the interleaved code words (*fig. 12*) and only a simple code is required to correct them. Note that the interleaving process does not involve adding more redundancy.

Interleaving enables errors occurring during the short-term fades due to multipath reception to be dispersed into the intervals between these fades. The short-term fades are spaced by a minimum of a half wavelength, 32.5 cm at 462 MHz, and a vehicle moving at the typical city speed of 30 km/h (19 mph) will transmit or receive 47 bits (at 1200 bits/s) while travelling this distance. Allowance must then be made for variation in fade separation, fade depth and vehicle velocity, so a useful interleaving period is likely to be several hundred bits. The longer-term fades due to shadowing lie further apart. It is impractical to combat the errors they cause with bit interleaving.

The benefit from interleaving is illustrated in *fig. 13*, which shows the success rate for a (31,16,3) random error-correcting code without interleaving, and interleaved over 992 bits (32 code words interleaved). By interleaving, the success rate at 10 dB mean signal level has been improved from 99.8% to 99.998% and the false rate (not plotted) will be improved by two orders of magnitude. A considerable improvement has therefore been obtained from interleaving. The random-access memory required to implement an interleaving period of 992 bits represents a practical size and it was found that further improvement in performance was only gradual as the interleaving period was increased beyond this.

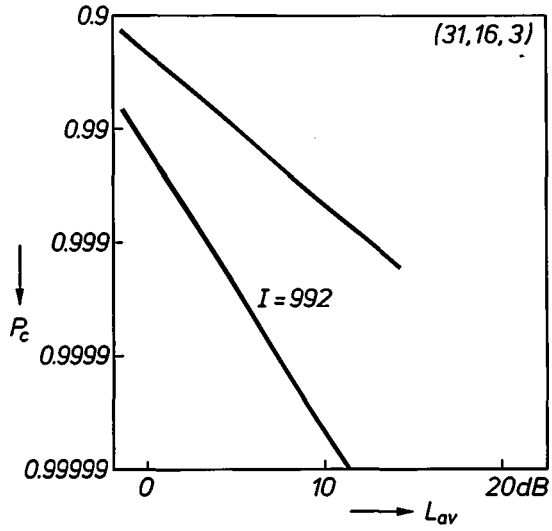
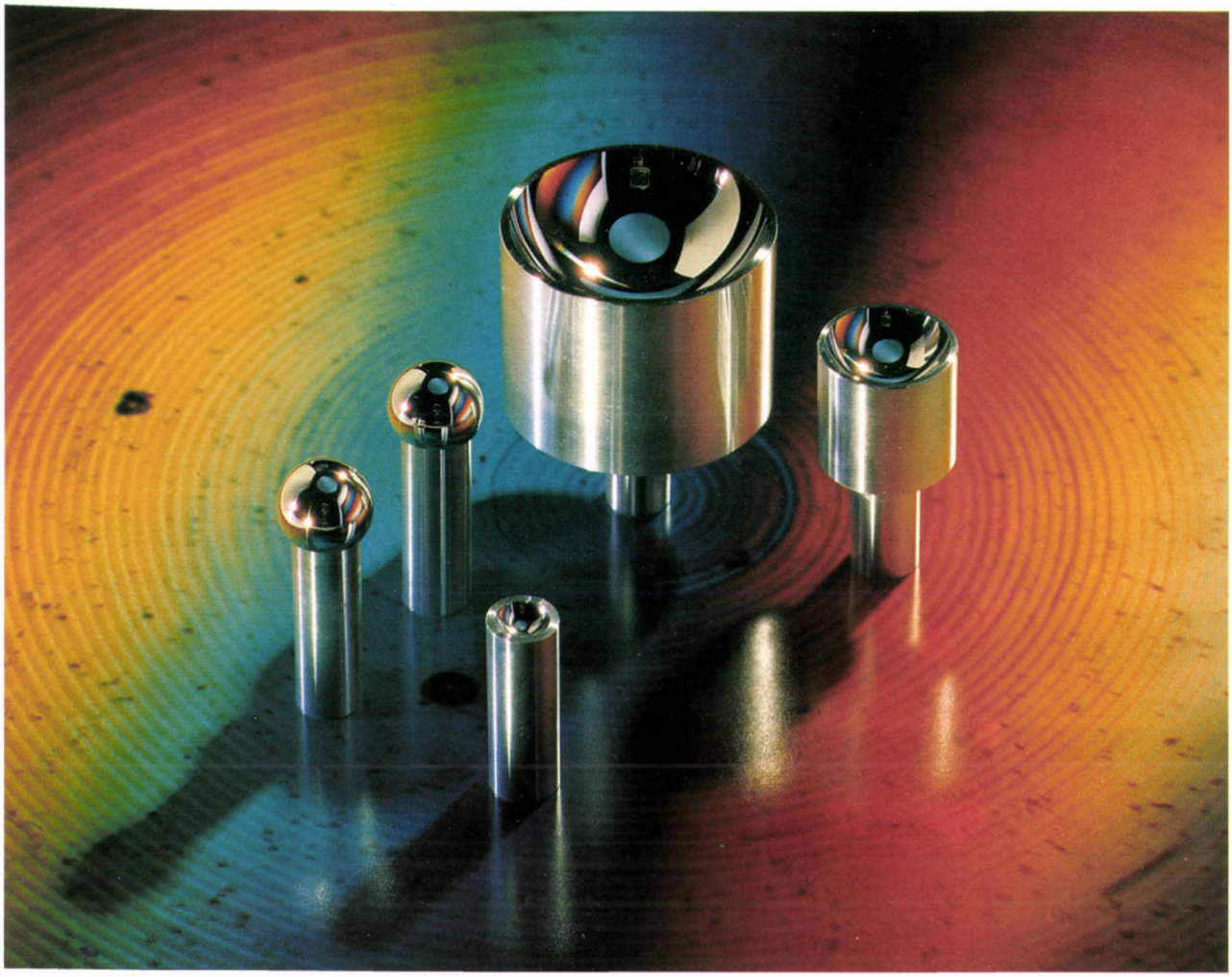


Fig. 13. Improvement du to bit interleaving. The success rate P_c of an error-correcting (31,16,3) code is increased considerably by interleaving over a period of $I = 992$ bits (32 code words).

Summary. Transmission of digital signals opens up new possibilities in the use of mobile radio, such as sending written messages in both directions or even pictures and maps. The level of the received signal is low, however, and very variable because of shadowing and fading; there is also considerable ignition noise from the surrounding traffic. During measurements made in a vehicle travelling in central London and near Philips Research Laboratories (PRL) at Redhill recordings were made of signal-level fluctuations and bit-error patterns. Bit-error rates were calculated from these recordings; the errors are often clustered. For reliable data transmission it is necessary to use an error-detecting or error-correcting code, in which check bits are added to the information bits. The collected statistical material has enabled the performance of various codes to be compared in the laboratory.



The background of this photograph is part of an enlargement (about 200×) of an aluminium surface machined to be 'optically smooth'; the original was made by means of Nomarski interference contrast. The concave or convex tops of the samples shown (maximum diameter 4 cm) are optically smooth surfaces, made on a numerically controlled high-precision lathe. The turning marks in the background have a depth of about 20 nm and a pitch of 10 μm. The black dots are relatively innocuous specks of impurities in the aluminium. The ability to produce perfectly reflecting curved surfaces, by diamond turning, has greatly extended the application of geometrical optics at Philips. The high-precision lathe developed for this type of work at Philips Research Laboratories is known as COLATH; it will shortly be described in this journal.

Grease-lubricated spiral-groove bearings

E. A. Muijderman, G. Remmers and L. P. M. Tielemans

Since the start of the sixties Philips Research Laboratories have been doing research on spiral-groove bearings [*]. The articles on this subject that have appeared in this journal have described the excellent characteristics of these bearings, especially when grease is used as the lubricant. Some successful professional applications have also been described. Large-scale application in consumer articles was not possible until quite recently, however, partly because of a large number of theoretical and practical aspects of spiral-groove bearings and the long time necessary for the development of an inexpensive method of manufacture and for testing important characteristics such as reliability and life and other special features. The use of these bearings in the recently introduced VR 2020 video cassette recorder, based on the new 'Video 2000' system, can therefore be considered as something of a landmark in the work on grease-lubricated spiral-groove bearings.

Introduction

A spiral-groove bearing is a self-acting bearing with a repetitive pattern of grooves in one of its two bearing surfaces. When these surfaces move in the right direction in relation to one another, the pumping action of the grooves produces an overpressure in the lubricant between them. Because of this overpressure the lubricant keeps the moving surfaces apart and a force is transferred. The first article on this subject in this journal [1] explained the principle of operation of the bearing and drew attention to its low friction and the possibility of making bearings with very small dimensions (e.g. with a diameter of only 3 mm).

In many applications in professional equipment the spiral-groove bearings are often lubricated with air. This does not contaminate the immediate surroundings of the bearing, it can operate at extremes of temperature and the speed of the shaft can be very high. In Britain and America gyrocompasses, flywheels, fans, small electric motors and gas circulators have been fitted with air-lubricated spiral-groove thrust and journal bearings. A number of interesting possible applications have also been investigated at Philips Research Laboratories; one of these is the 'push-pull' bearing, a thrust bearing with self-adjusting internal preloading [2].

Prof. Dr Ir E. A. Muijderman, Ing. G. Remmers and Ing. L. P. M. Tielemans are with Philips Research Laboratories, Eindhoven.

For many applications, and here we include almost all consumer articles, air lubrication is unsuitable, however, because the load-carrying capacity of the air-lubricated bearing is usually too small for the dimensions and speeds normally used in this equipment. To obtain bearings with sufficient load-carrying capacity we need a lubricant that has a much higher viscosity, such as oil or grease. Oil lubrication is used for example in the bearing system we have developed for ultracentrifuges; grease lubrication is used in the bearing system of the flywheel for stabilizing space vehicles described previously in this journal [3]. An advantage of greases that became clear at an early stage [1,4] is that they do not leak out of a bearing as readily as do oils. And even after long periods of use some greases retain their excellent lubricating properties so that the life of grease-lubricated spiral-groove bearings can be very long [4,5]. All ten of the flywheels are still operating perfectly after running for more than ten years in laboratory tests although their bearings have never been re-lubricated.

The potential advantages of grease lubrication and our favourable experience in professional applications led us to a study of grease-lubricated spiral-groove bearings for applications in consumer articles. This work has resulted in a number of suitable bearing designs. Further studies made in cooperation with grease manufacturers' laboratories have led to the

development of greases with appropriate properties for the special conditions in spiral-groove bearings. From the results obtained we conclude that these bearings now operate more reliably and more predictably than the porous oil-impregnated bearings of sintered metal ('sintered bearings') that are widely used in consumer products.

the choice of groove pattern and lubricant. This description will include results obtained from the research of the past few years [6–9]. We shall then cover in somewhat greater detail the use of greases and show that the exacting requirements for viscosity, rheological behaviour and stability over long operating periods can, to some extent, be met by

Fig. 1. The VR 2020, the new Philips video cassette recorder.

A very interesting application for spiral-groove bearings was found when a new generation of video cassette recorders (VCR) based on the 'Video 2000' system was being developed for domestic use. The desired playing time of 2×4 hours per cassette placed such exacting requirements on the running accuracy of the head disc that the sintered-bronze bearing system used in earlier types was no longer adequate for the head-disc motor. Working together with technologists and motor and VCR designers from the Audio and Video Divisions we have developed grease-lubricated spiral-groove bearings that do meet the requirements. These bearings are highly suitable for this particular consumer product because they combine excellent technical and practical features: they will locate the shaft very accurately, they will handle high loads, they are stable, have a constant friction and a long life, they are silent, they are lubricated for life, and they are very reliable and reasonably inexpensive. They are used with success in the VR 2020, the new recorder that has been on the market since last year (*fig. 1*).

In this article we shall describe first of all the general operation of spiral-groove bearings, including making a correct selection of lubricant composition. We shall then discuss manufacture of spiral-groove journal bearings, with emphasis on the selection of the bearing material and the way in which it is machined. Finally, we shall look at the application of grease-lubricated spiral-groove journal bearings in the head-disc motor of the VR 2020.

---

[*] The term 'spiral-groove bearing' in this article is used both for spiral-groove thrust bearings and for helical-groove journal bearings.

[1] E. A. Muijderman, Philips tech. Rev. **25**, 253, 1963/64.
See also by the same author:
Polytech. T. A **20**, 154A, 1965 (in Dutch);
Proc. Instn Mech. Engrs **180**, 174, 1965/66;
Scientific American **214**, March 1966, page 60.

[2] H. J. W. M. Volman, Philips tech. Rev. **35**, 11, 1975.

[3] J. P. Reinhoudt, Philips tech. Rev. **30**, 2, 1969.

[4] G. Remmers, Philips tech. Rev. **27**, 107, 1966.

[5] G. Remmers, Philips tech. Rev. **34**, 103, 1974.

[6] G. Remmers, in: Advances in tribology (papers, 1976 Convention, Durham), page 13; Instn Mech. Engrs, London 1978.

[7] J. Bootsma and L. P. M. Tielemans, Trans. ASME F (J. Lubr. Technol.) **99**, 215, 1977.

[8] J. P. Reinhoudt, On the stability of rotor-and-bearing systems and on the calculation of sliding bearings, Thesis, Eindhoven 1972; also published as Philips Res. Repts. Suppl. 1972, No. 1.

[9] J. Bootsma, Liquid-lubricated spiral-groove bearings, Thesis, Delft 1975; also published as Philips Res. Repts. Suppl. 1975, No. 7.

Operation of spiral-groove bearings

To illustrate the operation of the bearing let us first examine a simple model consisting of a plain bearing surface in motion and, parallel to it and underneath it, a stationary bearing surface with a pattern of parallel grooves at an angle α to the velocity vector; see *fig. 2*. The presence of the grooves produces an overpressure between the bearing surfaces because the lubricant, the 'viscous medium', is 'dragged' through narrowing gaps (the 'wedge effect'). Because the grooves are at an acute angle, lubricant is also pumped along the grooves so that the pressure increases in this direction. The mean pressure difference Δp between the two ends of the groove pattern ($y = d$ and $y = 0$) depends on a number of quantities

$$\Delta p = \frac{\eta v d g}{h_r^2},$$

where η is the viscosity of the lubricant, v the velocity of the upper bearing surface, h_r the distance between the upper bearing surface and the ridges in the lower bearing surface, and g is a function of the groove angle α, the ratio β of the ridge width b_r to the groove width b_g, and the ratio δ of h_r to the groove depth h_g. Calculations show that for grooves of rectangular cross-section the function g has a maximum value (about 0.09) when $\alpha \approx 16°$, $\beta = 1$ and $\delta \approx 0.4$.

In most cases both theory and practical applications relate to grooves of rectangular cross-section. This is not essential for the bearing action: any cross-section is in principle possible and can be used in practice provided it has been dimensioned correctly. In *fig. 3* the function g is plotted against δ for a triangular groove cross-section for different values of α. The envelope curves have also been drawn for rectangular, circular and triangular cross-sections. These curves show that the greatest pressure is produced with a rectangular cross-section and also that the pressure is not much lower for the other two cross-sections.

The results of the simple model with parallel grooves can be used to determine the best shape of groove for a thrust bearing [10]. When the shaft is rotating in the correct direction with respect to the groove pattern the grooves pump lubricant towards the centre of the bearing. This bearing can be thought of as being split up into a large number of elements each resembling the model described above; see *fig. 4*. Each 'ring' at an arbitrary distance r from the centre with a radial width Δr functions as a pump. The pumping action developing the pressure is at its most effective when the pressure difference across Δr is as large as possible. This means that all the grooves should be at the same angle to the local velocity vector. This requirement is satisfied if the grooves have the shape of a logarithmic spiral. Similarly it may be deduced that in journal bearings the grooves should take the form of a helix.

It will be clear that these pumps will produce a pressure in the lubricant between the rotating and stationary bearing surfaces. The load-carrying capacity of the bearing can be calculated by integrating the pressure over the entire bearing surface. Allowance must be made here for the finite number of grooves.

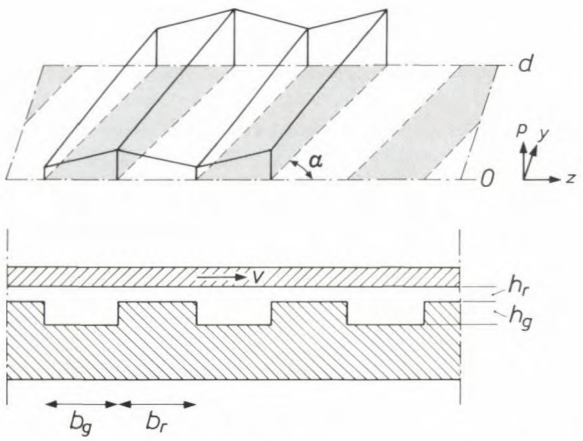


Fig. 2. A diagram showing the distribution of the pressure p in the lubricant, the 'viscous medium', between a plain bearing surface moving at a velocity v and a stationary bearing surface parallel to it, with grooves that have a rectangular cross-section and are at an angle α to the velocity vector. h_r height of gap above the ridges, h_g groove depth, b_r ridge width, b_g groove width. There is a constant pressure gradient in the direction of the grooves.

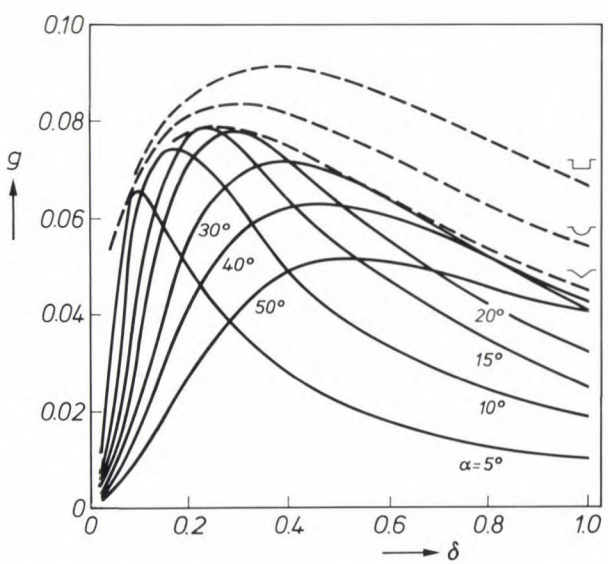


Fig. 3. The function g for $\beta = 1$ plotted against δ for different values of α. The solid lines refer to grooves of triangular cross-section. The dashed lines represent the envelope curves for rectangular, circular and triangular groove cross-sections. The highest value (about 0.09) is obtained for grooves of rectangular cross-section when $\alpha \approx 16°$ and $\delta \approx 0.4$.

The three-dimensional representation of the pressure distribution in a spiral-groove thrust bearing that we gave in an earlier issue of this journal [11] applies only to a model with an infinite number of grooves. In an actual practical bearing the pressure ripples that we have shown in fig. 2 appear on the outside of this 'pressure hill' as a kind of modulation. The pattern of the pressure distribution in a thrust bearing with k grooves is shaped like an inverted pudding basin with k ribs on its surface.

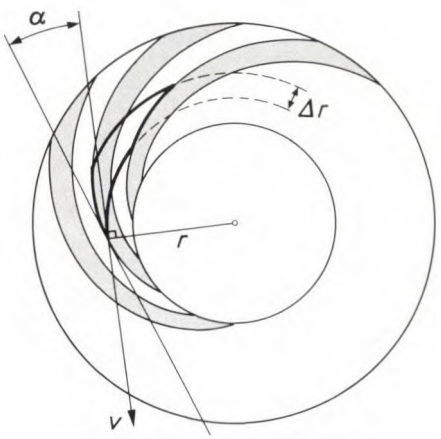


Fig. 4. Part of a spiral-groove pattern for a thrust bearing. Each 'ring' of width Δr, at a distance r from the centre, pumps lubricant towards the centre. The pumping action is strongest when the grooves have the shape of a logarithmic spiral, so that the angle α between the local velocity vector v and the tangent to the groove is constant over the entire grooved surface.

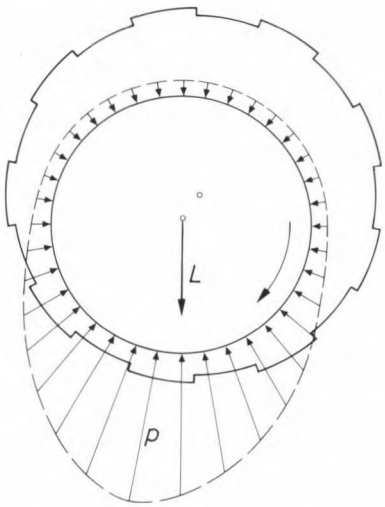


Fig. 5. Diagram of the distribution of the pressure p in a spiral-groove journal bearing in which the shaft runs eccentrically because of the load. L load vector. Only the overpressure in the constricted part of the bearing gap gives the bearing a radial load-carrying capacity. The detailed effects of the grooves on the pressure distribution are not shown.

The grooves in a spiral-groove journal bearing also build up a pressure between the bearing surfaces. However, when the shaft is concentric with the bearing bush this does not produce a radial load-carrying capacity. Only if the shaft is in an eccentric position is a radial load-carrying capacity produced by the overpressure created in the narrowing part of the bearing gap; see *fig. 5*. The grooves ensure that the lubricant is kept inside the bearing.

The groove pattern

Fig. 6 gives diagrams of three different groove patterns for a spiral-groove thrust bearing. For a 'blind' thrust bearing a completely grooved bearing surface is used (fig. 6*a*). In this bearing there is no net flow of lubricant from the outside to the centre. If the same pattern is used for a thrust bearing with a straight-through shaft, the pumping action cannot maintain a sufficiently high pressure because the lubricant leaks away too much in the immediate vicinity of the shaft. This leakage can be restricted by reducing the radial length of the groove pattern (fig. 6*b*); the plain part then acts as a resistance to the flow. If a 'herringbone pattern' (fig. 6*c*) is used, the leakage around a straight-through shaft can be eliminated completely because the pumping actions of the two sections of the herringbone are in opposing directions.

A herringbone pattern is also suitable for journal bearings. The groove pattern can extend over the entire length of the bearing or merely over a part of it, with a part of the bearing surface between the two sections of the herringbone left plain; see *fig. 7*. The two sections of the herringbone do not have to be of the same length. When grease is used as the lubricant, an asymmetrical herringbone pattern with a completely grooved bearing surface is preferred. An asymmetrical pattern of this kind ensures that any spare grease remains on the side of the bearing that the designer intended, because the longest section of the herringbone can deliver a higher pump pressure than the shortest. When the bearing surface is completely grooved it is easier to prevent the occurrence of underpressure areas in the grease layer that might cause cavitation. Air bubbles can accumulate in such areas, so that the space between the bearing surfaces becomes incompletely filled with grease [12].

For the operation of a spiral-groove bearing it does not matter whether the groove pattern is in the rotating bearing surface or in the stationary one. It

[10] E. A. Muijderman, Spiral groove bearings, Thesis, Delft 1964; also published as Philips Res. Repts. Suppl. 1964, No. 2.
[11] See fig. 21 of the first article by E. A. Muijderman referred to in note [1] (Philips tech. Rev. **25**, 253, 1963/64).
[12] J. G. G. Bos, Philips tech. Rev. **35**, 137, 1975.

does make a difference, however, whether the pattern rotates with respect to the load vector or not. *Fig. 8* shows four different situations for a journal bearing with a non-rotating load vector. In two cases the grooved surface rotates and in the other two the plain surface rotates. Situations where the load vector rotates are also possible, and their behaviour can also be predicted theoretically.

Also, air offers no protection from corrosion or from damage when the bearing surfaces touch on starting and stopping. For bearings that do not operate at such high speeds, oil is therefore the usual lubricant. A disadvantage of oil is that it can easily leak away when the bearing is at rest, because of the action of gravity. Extra precautions then have to be taken to maintain continuous lubrication and prevent the environment of the bearing from being contaminated. We took the unusual step of selecting *grease* as the lubricant instead of oil. So far, the spiral-groove bearing is the only self-acting bearing that permits this option without the need for re-lubrication. When grease is used, much less leakage occurs at rest, because the grease does not start to flow until the shear stress exerted on it exceeds a certain limiting value, the yield point. In an operating spiral-groove bearing a grease behaves almost like an oil.

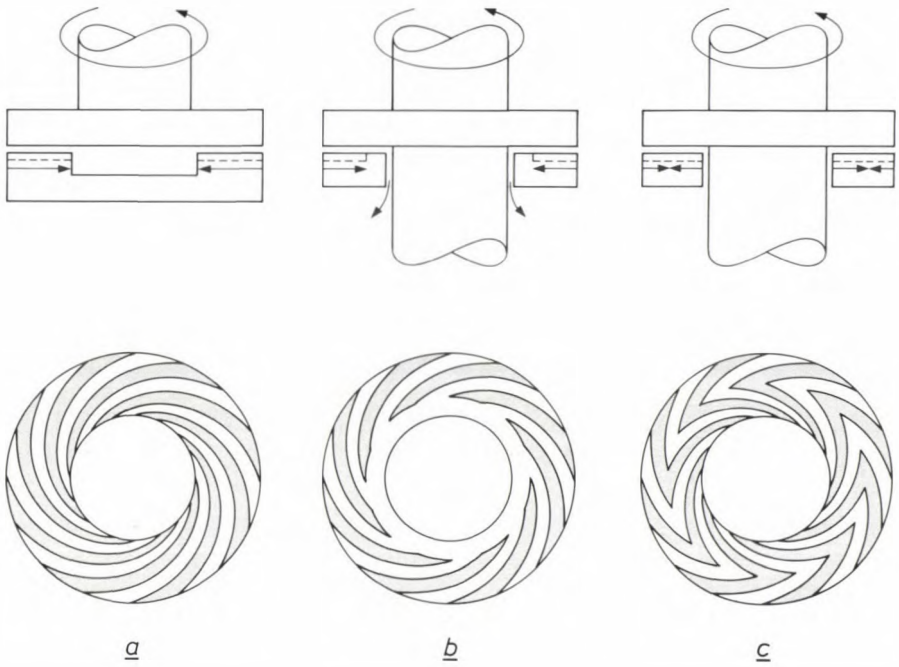


Fig. 6. Three types of spiral-groove bearing for a thrust-bearing system. *a*) Completely grooved blind thrust bearing. *b*) Bearing for a straight-through shaft, with shorter grooves. *c*) Herringbone bearing for a straight-through shaft. The straight arrows indicate the direction in which the lubricant is pumped. The downward-pointing arrows indicate the places where the lubricant leaks out.

The 'viscous medium'

The use of air as a lubricant has the advantage that the bearing is always 'immersed' in a bath of lubricant. Air-lubricated spiral-groove bearings are highly suitable for use at high temperatures and very high speeds [1]. A disadvantage of air is its low viscosity; this often makes it necessary to use bearings with impracticably large diameters and a very small clearance.

Fig. 7. Herringbone pattern for a completely grooved bearing surface (*left*) and for a partially grooved bearing surface (*right*) in a spiral-groove journal bearing. The two groove patterns are symmetrical: the two sections of the herringbone that pump lubricant to the centre are of the same length. When grease is used as the lubricant an asymmetrical pattern is preferred.

The operation of the bearing is affected by flow movements of the lubricant. We can consider the net flow through the bearing ('macro-flow'), or the flow in the grooves and above the ridges ('micro-flow').

If sufficient lubricant is fed into the bearing from outside, a journal bearing with an asymmetrical groove pattern will itself create macro-flow because one section of the herringbone has a stronger

pumping action than the other. The occurrence of macro-flow is sometimes desirable since it may help to cool the bearing. For some bearing geometries and operating conditions this macro-flow may be stronger than the cooling flow produced by an external pressure difference. Macro-flow has practically no effect on the load-carrying capacity or the stiffness of a journal bearing, but does have an effect on the stability and the frequency of the 'half-speed whirl' generated by the bearing [13].

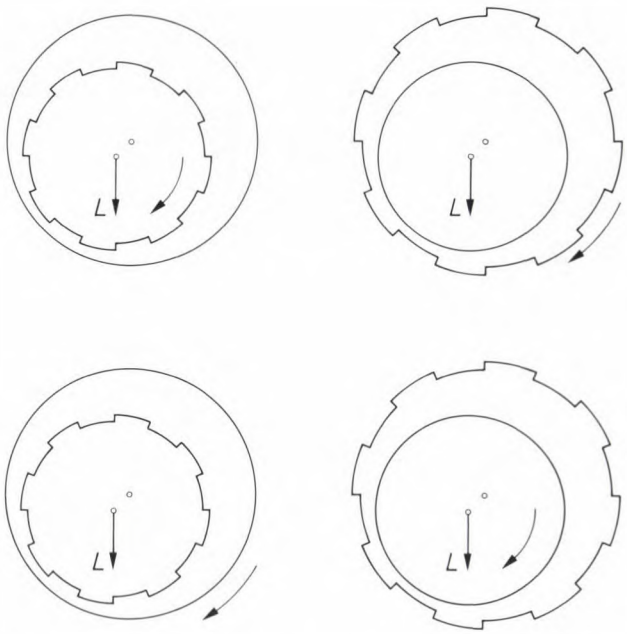


Fig. 8. Diagram showing four situations for a journal bearing with a non-rotating load vector L; grooves in shaft or bearing bush; shaft or bearing bush rotates. The bearing action depends only on whether the grooves rotate with respect to the load vector (the upper two situations) or are stationary with respect to the load vector (the lower two situations).

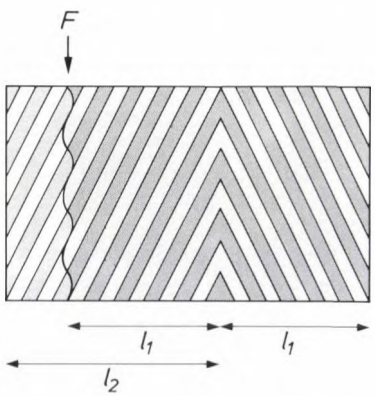


Fig. 9. Asymmetrical herringbone pattern of a completely grooved bearing surface in an oil- or grease-lubricated spiral-groove journal bearing. When there is a limited amount of lubricant a free boundary F arises in the left-hand section (length l_2) at a distance from the centre that is equal to the length l_1 of the right-hand section so that the two sections pump equally strongly towards the centre.

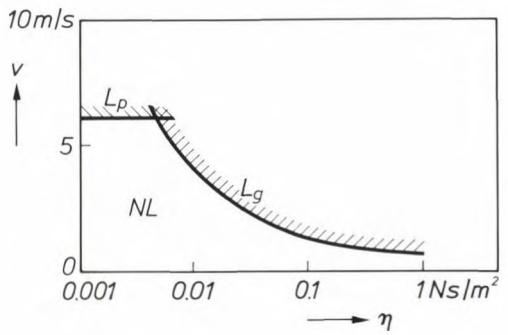


Fig. 10. Diagram showing leakage of lubricant in a spiral-groove journal bearing as a function of the circumferential velocity v and the viscosity η of the lubricant. *NL* no leakage. L_p leakage via the plain bearing surface; occurs only at a high circumferential velocity if the lubricant has a low viscosity. L_g leakage via the grooved bearing surface; occurs even at a low circumferential velocity if the viscosity is high.

The macro-flow that arises, after starting, in a journal bearing with an asymmetrical groove pattern and with a limited quantity of lubricant stops as soon as the two sections of the herringbone have been filled to the same extent; see *fig. 9*. This means that the longer section will now have a 'free boundary', which acts as a non-leaking seal because of the surface tension, provided the circumferential velocity is not too high. The velocity at which leakage commences will depend to a large extent on the viscosity of the lubricant [7]; see *fig. 10*. If the viscosity is low, leakage in most spiral-groove journal bearings will take place only at high speeds at the plain bearing surface, whereas it will occur much sooner along the grooved bearing surface if the viscosity is high.

Even when there is no macro-flow, there will still be micro-flow in a spiral-groove bearing. The complex flow field present in the groove pattern causes a quantity of lubricant to flow over the ridges towards the outer edge, in other words *out* of the bearing. The same quantity of lubricant flows back from the outer edge through the grooves *into* the bearing.

Greases for spiral-groove bearings

A grease for a spiral-groove bearing must have the correct viscosity and yield point. The existence of a yield point τ_y implies not only that the grease will not

[13] In 'half-speed whirl' the centre-line of the shaft rotates about the centre-line of the bearing bore at an angular velocity that is approximately half the angular velocity of the shaft. In certain situations the amplitude of the whirl gradually increases, causing the shaft to come into contact with the bearing bush. See also, for example, the articles by E. A. Muijderman [1,10], J. P. Reinhoudt [3,8], and J. Bootsma [9].

begin to flow until the shear stress τ exerted upon it is greater than τ_y, but also that the dynamic viscosity η depends on the shear rate $\dot{s}$:

$$\eta = \tau/\dot{s} = \tau_y/\dot{s} + \eta_\infty .$$

The viscosity only approaches a constant value η_∞ at high shear rates. Strictly speaking it would be more appropriate here to speak of 'the apparent dynamic viscosity', but for convenience we shall continue to talk about 'the viscosity'. The diagram of *fig. 11* illustrates the difference between a fluid with a yield point (a Bingham fluid) and a Newtonian fluid, which does not have a yield point and has a viscosity independent of the shear rate.

As well as meeting the requirements for viscosity and yield point a suitable grease must also satisfy some other general requirements. It must, for instance,

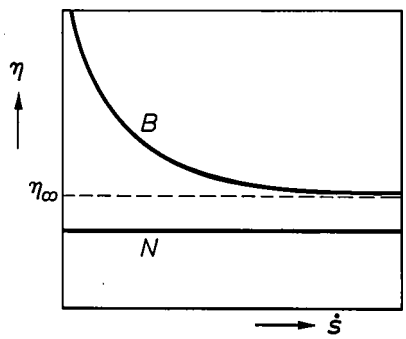


Fig. 11. Viscosity η as a function of the shear rate $\dot{s}$. For a Newtonian fluid *N*, η is constant; for a Bingham fluid *B*, η is a decreasing function of $\dot{s}$, which only tends to a constant value η_∞ when the shear rate is high.

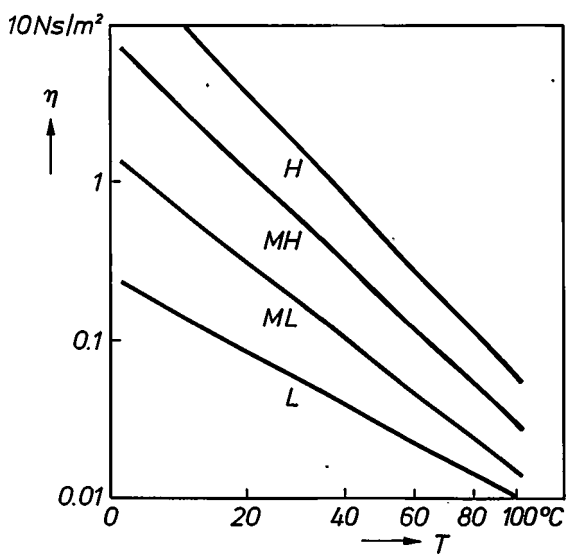


Fig. 12. Viscosity η of four basic greases that we used, as a function of temperature *T*. The designations *L*, *ML*, *MH* and *H* stand for 'light', 'medium-light', 'medium-heavy' and 'heavy'. This series of greases will cover a wide viscosity range.

protect the bearing surfaces from any damage that might occur when they are in contact with one another, even during rotation, and from corrosion from outside, but it must not attack the bearing surfaces or other structural materials. In many consumer articles it is extremely difficult or even impossible to relubricate the bearing once it has been assembled. The grease must therefore have a long life. This means that it must be mechanically and chemically stable, must not oxidize, and that the base oil must not evaporate or leak away from the grease.

Because of the special conditions in spiral-groove bearings it is difficult for a grease to meet all the above requirements in full. It must, for instance, have a sufficiently high yield point to prevent leakage when at rest, yet all of the grease must start to flow as soon as rotation starts. The amount of grease used in a spiral-groove bearing is usually very small, only a few tens of mm^3 per bearing. The grease is continuously subject to shear; the shear rate in the bearing gap is about 10^4 to 10^5 s^{-1}. We should remember here that a required life of several thousand hours is more the rule than the exception. There is also an intensive circulation of grease (micro-flow) in the bearing so that there is always 'fresh grease' arriving at the edges of the bearing and being exposed to oxygen there. Because of these special circumstances greases for spiral-groove bearings should generally be quite different in composition from greases used in ball bearings, for example.

Composition

A grease consists mainly of a base oil and a thickener; the first functions as the fluid lubricant, and the second determines the consistency.

In simple terms we could say that the thickener retains the liquid lubricant like a sponge. Another concept is to imagine the thickener as a three-dimensional network resembling a bundle of sticks to which the base oil is attached by physical and chemical bonds. When the shear stresses are small, the 'sticks' of the thickener deform elastically; the grease behaves like an elastic solid. When the stresses exceed the yield point, the 'sticks' slide about on one another; the grease then behaves like a plastic material. The sliding sticks break and begin to line up with the direction of flow so that the mixture of sticks and oil begins to behave more and more like a fluid. As the shear stress increases, the pieces of broken stick become smaller, until equilibrium is reached. The grease then begins to approximate more closely to a real fluid containing solid round particles. If the shear is stopped after a while, the thickener particles start to reorientate themselves somewhat and to recrystallize, but the yield point now has a considerably lower value than in the unworked grease.

To cover a sufficiently large range of viscosity it is not in fact necessary to stock greases of many different viscosities; in our view a range of at least four

should be sufficient. The four greases should have the same type of base oil and the same thickener, so that intermediate viscosities can be obtained by mixing. The differences in viscosity then depend mainly on differences in the mean molecular weight of the base oil. *Fig. 12* shows the viscosity as a function of temperature for a range of greases that we use. The greases have been designated *L* ('light'), *ML* ('medium-light'), *MH* ('medium-heavy') and *H* ('heavy'). The relation between the temperature and the viscosity can be expressed as:

$$\eta(T) \approx k \exp [b/(T + 95)],$$

where k and b are constants for a given grease, and T is the temperature in degrees Celsius. The variation with temperature increases with the mean molecular weight of the base oil. For a medium-heavy oil a rise in temperature of 10 °C reduces the viscosity by almost a half. Because it contains a relatively small amount of thickener, a suitable grease has a viscosity not much higher than that of its base oil: usually within a factor of two. The temperature dependence of the viscosity of a grease is always less than that of its base oil: as temperature increases, more base oil combines with the thickener particles or becomes attached to them to give 'thermal swelling', and the viscosity therefore decreases less sharply.

Rheological behaviour

Because re-lubrication is undesirable or sometimes even impossible the supply of grease to the bearing surfaces must be assured. In a spiral-groove journal bearing the grease is introduced into a reservoir immediately after assembly of the bearing at a point that allows it to come into contact with the wall of the reservoir and with the wall of the shaft. When the bearing runs for the first time, the grease has to be transported from the reservoir to the bearing gap. It is then forced further into the gap by the pumping action of the grooves. The distribution of the grease can be observed by using a transparent bearing, for example of polymethyl methacrylate (Perspex); see *fig. 13*. Greases that are easily transported to the bearing gap are said to have good 'slumpability'. Such greases require an accurate match between the consistency and the elasticity, quantities which are dependent on the thickener content and the presence of certain polymers.

As we have already stated, we prefer an asymmetrical groove pattern for journal bearings so that a reserve of grease can be stored at a predetermined location. This means that there is a free boundary in the bearing gap (fig. 9). The spare grease compensates for any loss of lubricant; this can happen, even with greases, because of effects like creep and evaporation. The life of the grease is also extended if the mechanical and chemical stresses are distributed over a relatively large amount of the grease. One of the requirements for the spare grease is that it can be easily transported to the bearing gap; it must not jell in the reservoir. The chance of this happening is considerable because the shear rate in the reservoir is about 100 times less than in the bearing gap. In other words the grease is 'stirred' much less intensively there. The jelling properties of the grease depend to a large extent on the type and content of thickener used and may also be affected by additives.

While, on the one hand, a grease for use in spiral-groove bearings must have good 'slumpability' and must not jell, its yield point must on the other hand always remain high enough to prevent leakage at rest.

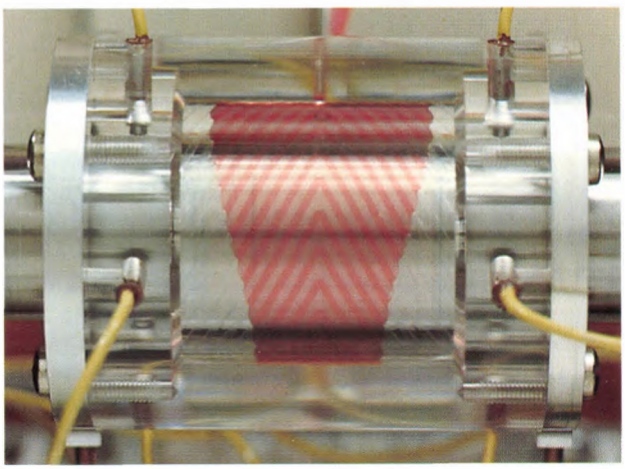


Fig. 13. Photograph of a test rig with a loaded spiral-groove journal bearing with a transparent Perspex bearing bush. The lubricant (in this case an oil) and the boundaries between the lubricant and the air (free boundaries) are clearly distinguishable.

This requirement is particularly difficult to meet because the thickener particles in the grease are broken up into even smaller pieces when the bearing is operating. When the bearing stops running, the grease is still warm and thin. It can therefore flow to the unfilled part of the bearing gap because of surface tension ('capillary creep') and it may leak downwards because of gravity. Both types of flow have to be prevented by a sufficiently high yield point. To prevent capillary creep the yield point τ_y must satisfy the condition

$$\tau_y > \gamma/B,$$

where γ is the surface tension (about 3×10^{-2} N/m) and B the length of the filled part of the bearing. In the vertically mounted spiral-groove journal bearing of Perspex that we used for observing creep and leak-

age effects this length is 19 mm. To prevent leakage due to gravity it is necessary that:

$$\tau_y > \frac{g\varrho(\Delta R + h_g)}{2},$$

where g is the acceleration due to gravity (about 10 m/s^2), ϱ is the density (about 1 g/cm^3), ΔR the clearance (about 30 μm), and h_g the groove depth (about 60 μm). Substituting the values for γ, B, g, ϱ, ΔR and h_g shows that when the yield point decreases during the use of this bearing, capillary creep appears first (at about 1.5 N/m^2), and that when the yield point then falls to about 0.5 N/m^2, gravity leakage will start.

Fig. 14 shows the qualitative variation in the creep as a function of the time of use for four different greases. In all four it takes some time before the yield point has fallen to such an extent that creep starts. For greases A, C and D the yield point continues to fall and, after a while full creep is attained: the bearing gap is then filled with grease over its entire length. This is not the case with grease B: after about 1500 hours there is hardly any further increase in the creep. It would seem that the size and the physical behaviour of the thickener particles in this grease have by then attained a stable final state. The grease B has a yield point that is ideal for spiral-groove bearings because it ultimately becomes constant at a sufficient magnitude.

If the content of certain polymers in a grease is too high this can lead to an undesirable effect that may increase the chance of leakage considerably: the spare grease may start to 'shuttle'. This effect, which as far as we know has not been observed before, can be clearly seen in an unloaded horizontally mounted transparent bearing with two groove patterns with about the same pumping action and a grease reservoir on either side. The spare grease, which we shall assume to be in the right-hand reservoir at first, starts to flow towards the left-hand reservoir. This flow continues until the right-hand reservoir is completely empty and a free boundary has been formed in the bearing gap. After a little while the process repeats itself, but in the reverse direction. This can go on for hundreds of hours. We believe that this shuttling movement is due to the 'micro-flow' mentioned earlier and to the effects of the shear rate and time on the viscosity. As a result of the micro-flow 'fresh' grease from the reservoir is exchanged for 'old' grease from the bearing gap. It may be assumed that it will take some time before the viscosity of the fresh grease has adapted to the much higher shear rate (it is about 100 times higher) in the bearing gap. Where delay times are long, this adaptation will be associated with a growing disturbance in the pressure equilibrium in the bearing gap, which makes the grease start to flow in the direction of the empty reservoir.

Changes in viscosity during the period of use

When a grease-lubricated spiral-groove bearing is in operation, there is an initial decrease in the viscosity as a result of the breakdown of the thickener. The measured decreases in viscosity can be described as a function of time t, at a constant temperature T, with sufficient accuracy by:

$$\eta(t,T) \approx \eta_0(T)[c_1 + c_2 \exp(-c_3 t)],$$

where η_0 is the viscosity of the base oil and c_1, c_2 and c_3 are constants. After a certain time the viscosity assumes an almost constant value because thickener breakdown stops or has no further effect on the viscosity. The motor power, which has to be large enough to overcome the higher initial viscosity, is therefore greater than necessary for the required load-carrying capacity and stiffness of the operating bearing. To keep this excess in motor power to a minimum, the decrease in the viscosity must be as small as possible. In *fig. 15* the viscosity is shown as a function

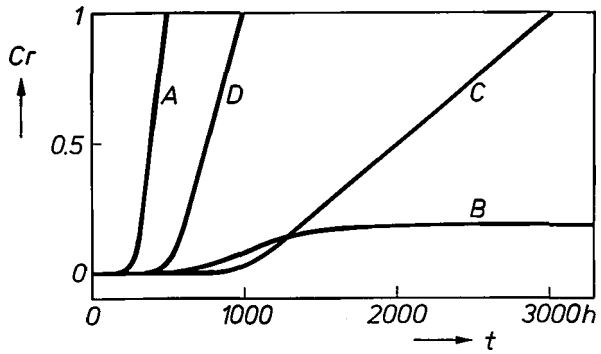


Fig. 14. Qualitative variation of the creep Cr as a function of the test time t for four greases in a spiral-groove journal bearing. $Cr = 0$: no creep, $Cr = 1$: full creep, with the longest section of the asymmetrical herringbone pattern also completely filled with grease. Eventually full creep occurs with greases A, C and D, but not with B.

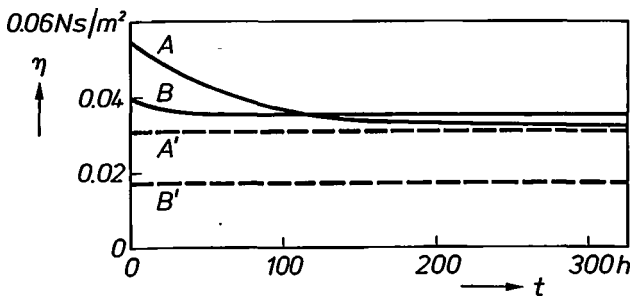


Fig. 15. Viscosity η as a function of the operating time t for greases A and B and their base oils A' and B' at a constant temperature of 55 °C and a constant shear rate of about 5×10^4 s^{-1}. B reveals a much smaller drop in viscosity than A: the viscosity remains considerably higher than that of its base oil B'.

of the operating time for two of the greases in fig. 14, *A* and *B*. The figure also gives the viscosity of the base oil for each of the two greases. In about 300 hours the viscosity of grease *A* falls from 0.054 Ns/m^2 to a constant value of 0.032 Ns/m^2, a fall of more than 40%. The viscosity of the grease is then only a few per cent higher than that of its base oil so that there must have been extensive breakdown of the thickener. The viscosity of grease *B*, on the other hand, decreases within some 50 hours from 0.040 Ns/m^2 to a constant value of 0.036 Ns/m^2, a fall of only 10%. In this case the viscosity is still more than twice as high as that of its base oil.

Effect on the life of spiral-groove bearings

The life of a grease-lubricated spiral-groove bearing is limited by a number of factors. The operation of the bearing may start to deteriorate if grease leaks out of the bearing or spare grease jells. Other factors affecting the life are the oxidation of the grease and the evaporation of the base oil. After a time these factors may cause the viscosity to increase considerably, so that the hydrodynamic frictional torque becomes too high. In *fig. 16* the frictional torque is shown as a function of operating time for the greases *A*, *B*, *C* and *D*. Grease *A* is known to be very resistant to oxidation. The frictional torque therefore remains constant after more than 10 000 hours. With grease *B* there is a noticeable rise in the torque after about 1000 hours, mainly due to creep and evaporation of the base oil. With grease *C* oxidation causes an increase in the frictional torque practically from the start. Grease *D* is very similar in composition to grease *C*, except that materials have been added to make the grease oxidation-resistant. After about 5000 hours the frictional torque has still not increased.

When a grease that oxidizes easily is used the materials of the shaft and bearing bush can also have a significant effect on the life; *Table I* illustrates an example of this. This effect is almost always the result of the catalytic effect of copper on the oxidation of greases.

Table I. Example illustrating how the materials of the shaft and bearing bush affect the life of grease-lubricated spiral-groove journal bearings. The results were obtained with a grease that was not very resistant to oxidation. When other greases are used the order may be different and the lives much longer.

Shaft material	Bearing-bush material	Life (hours)
bronze	Perspex	200
bronze	steel	250
glass	Perspex	950
glass	steel	>1000

In finally choosing the grease to be used, a compromise must be found between the minimum creep (fig. 14), the lowest possible fall in viscosity at the start (fig. 15) and the smallest possible increase in the frictional torque during long periods of use (fig. 16). Because of its low creep and low fall in viscosity we chose grease *B*; we are able to reduce the increase in the frictional torque by making a slight change in the composition of the base oil.

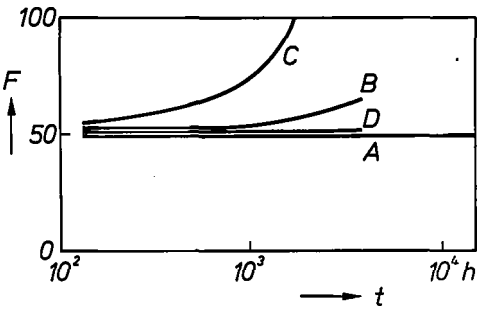


Fig. 16. Frictional torque *F* (in arbitrary units) as a function of the operating time *t* of a spiral-groove journal bearing lubricated by one of the greases shown in fig. 14. The rise in the frictional torque when *B* or *C* is used is the result of an increase in the viscosity. The main reason for this increase for *B* is loss of base oil and for *C* it is oxidation.

Manufacture of spiral-groove journal bearings

Bearing material

Either metal or plastic can be used as the bearing material. For metal spiral-groove bearings we prefer bearing bronze because this is a material that gives good dry-running behaviour: if the bearing surfaces come into contact (e.g. on starting or stopping) hardly any damage occurs. The material for plastic bearings must also have good dimensional stability; polyacetals seem to be very suitable.

One of the main disadvantages of plastics is that their thermal conductivity is much worse than that of metals. If the heat generated in the layer of grease is dissipated mainly by way of the bearing bush — and not by way of the metal shaft — the grease may become too hot and hence its viscosity too low, so that the load-carrying capacity and the stability of the bearing may become too small. In addition, the thermal expansion of plastics is so much greater than that of metals that, as the temperature increases, the expansion of the bearing clearance also reduces the load-carrying capacity and the stability. However, good use can be made of this relatively high thermal expansion, for example by enclosing a plastic bearing

bush in a metal sleeve; see *fig. 17*. Because of the difference in expansion coefficient and modulus of elasticity the bearing gap now becomes smaller instead of larger as the temperature rises, and this compensates for the effects of the fall in viscosity.

When a load is applied, plastics give a greater elastic deformation than metals. Plastics can also undergo plastic deformation as a result of creep

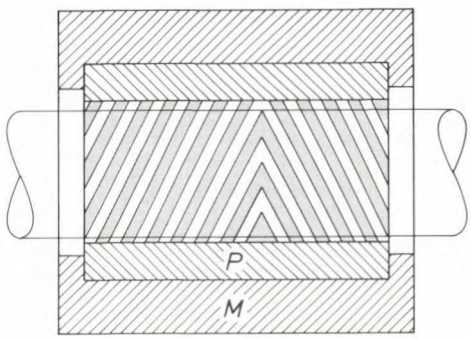


Fig. 17. Diagram of a spiral-groove journal bearing with plastic bearing bush *P* enclosed in a metal sleeve *M*. If the plastic wall is thick enough the bearing gap may decrease instead of increase with temperature.

effects. On the other hand, plastics have a greater internal attenuation than metals and therefore transmit vibration and sound less readily between shaft and frame. In equipment in which mechanical vibration must be kept to a minimum (e.g. record players) or where the sound level has to be as low as possible (e.g. electric shavers) this can be an advantage.

The poor electrical conductivity of plastics compared with metals can be both an advantage and a disadvantage. Bearings made of plastic are safer because they are electrically insulating, but metal bearings have the advantage that the suppression of electrical interference is easier.

An advantage of plastics is that they have hardly any effect on the oxidation of greases, so that grease-lubricated plastic bearings can have a relatively long life. A practical advantage is that the bearings can be given a colour to indicate characteristics that are not immediately obvious such as the groove dimensions or the groove pattern. A practical disadvantage with plastic bearings is that it has not so far been easy to make them with the same dimensional and geometrical accuracy as metal bearings. The result of this is that greater differences in load-carrying capacity and friction will occur in products with plastic bearings for the present than in products with metal bearings. This may mean, for example, that greater attention has to be paid to speed-control systems.

The advantages and disadvantages we have mentioned are so varied that for each particular application it is necessary to consider whether metal or plastic bearings are the more suitable. Nevertheless we do expect that when more experience has been gained with the manufacture and application of spiral-groove bearings, plastic bearings will be increasingly used.

Machining

The two most important machining operations in the manufacture of spiral-groove journal bearings are the production of the bearing bore and the production of the grooves. The bore must be very accurate in dimensions and shape (to within a few μm). The most important requirement for the groove pattern is that the variation in groove depth should be small (also a few μm).

Conventional metal-removal operations can be used for making metal spiral-groove bearings. Both the bearing bore and the grooves can be machined on precision lathes. It is also possible to turn the bearing bore only and to cut the grooves using a tapping process similar to that used for making threaded holes for screws. Particularly accurate bores can be made by broaching. A broach is a tool with an all-round cutting action: a shaft with a 360° cutting edge is pulled or pushed through a cylindrical hole.

In addition to these metal-removal machining operations there are also forming processes based on plastic deformation. Certain soft and ductile metals can be made to flow around a mandrel by applying a pressure on all sides. Once the pressure has been removed, the formed product springs elastically outwards and can be removed from the mandrel. Plastic deformation may also be imparted by means of rolling balls. Grooves can for example be made in an undersized bore by pressure exerted by rolling balls supported by a cylindrical core. The bore has still to be finish-turned or broached to remove the raised

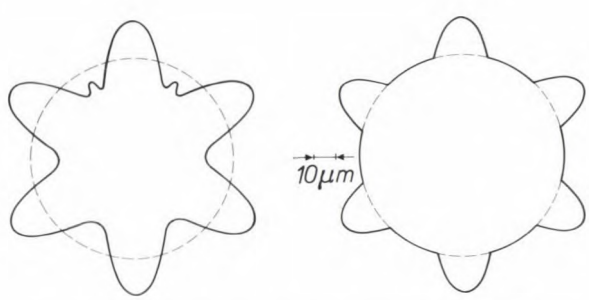


Fig. 18. Results of roundness measurements ('Talyround' measurements) on a bearing bore with rolled-in grooves. *Left:* not finish-turned. *Right:* finish-turned so that the raised portions have disappeared. The circles denote the original dimension of the bearing bore before the grooves are rolled in.

portions. *Fig. 18* shows the results of roundness measurements ('Talyround' measurements) on a bearing bore with rolled-in grooves before and after finish-machining.

Physical and chemical metal-removal processes such as spark machining [14] and all kinds of etching operations are less attractive on the whole than mechanical processes for the manufacture of spiral-groove journal bearings.

Injection moulding of thermoplastics is a fast and inexpensive way of producing plastic spiral-groove journal bearings. The injection-moulding process has to be performed under rigidly specified conditions. The temperatures in the mould must be very carefully controlled to obtain the optimum reproducibility. In injection moulding the grooves are made at the same time as the bearing bore. This introduces a complication, since a ridged mould plug has to be removed from a grooved bearing bore.

The spiral-groove bearings in the VR 2020 video cassette recorder

Let us now look more closely at the application of grease-lubricated spiral-groove journal bearings in the VR 2020 video cassette recorder (fig. 1). *Fig. 19* illustrates the motor that drives the disc carrying the two magnetic heads in the VR 2020 [15]. As we indicated briefly in the introduction, the reason for using spiral-groove journal bearings is closely associated with the demand for a longer video-cassette playing time and hence a higher information density on the tape, which in turn requires that the head disc should have greater true-running accuracy. For the Philips first-generation video cassette home recorders, the N 1500, the signal track-width was 130 μm and the spacing between tracks 57 μm. At a tape speed of 14.29 cm/s this gave a maximum playing time of one hour. In the next generation, the N 1700, the track-width was reduced to 85 μm and no space at all was left between the tracks. At a tape speed of 6.56 cm/s this gave a maximum playing time of two and a half hours. The transition from the N 1500 to the N 1700 did not affect the bearing system for the head disc: the required true-running accuracy for the head disc was still possible with a sintered-bronze bearing system [16].

Sintered-bronze bearings, however, were not suitable in the 'Video 2000' recorders, which have a track-width of only 22.5 μm (with zero spacing) and a maximum playing time of 2×4 hours at a tape speed of 2.44 cm/s. The use of expensive precision ball bearings is not without problems. It has been found that grease-lubricated spiral-groove journal bearings do meet the demands made of the bearing system in the head-disc motor.

The head-disc motor of the VR 2020 has a speed of 1500 rev/min. The shaft is mounted vertically, and supported in two spiral-groove journal bearings; see *fig. 20*. The bearings take up the radial load due mainly to the tape tension. This load is about 1.75 N for the upper bearing and about 0.75 N for the lower bearing. Because of possible unbalance in rotor or head disc, allowance has also to be made for an unbalance force: a maximum of 0.3 N for the upper bearing and a maximum of 0.2 N for the lower bearing.

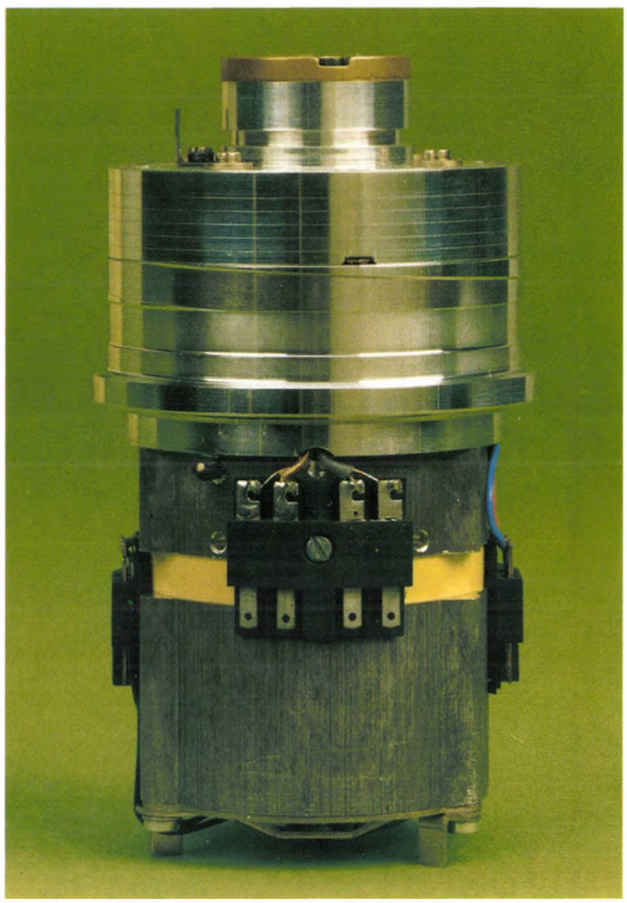


Fig. 19. The head-disc motor of the VR 2020. The photograph shows only one of the two magnetic heads. Below this can be seen the angled guide for the tape.

The diameter of the shaft and the outside diameter

[14] Spark machining is a processing method in which small particles of material are removed by means of a spark discharge. See for example C. van Osenbruggen, Philips tech. Rev. **30**, 195, 1969.

[15] See for example fig. 39 in the review article on modulation by F. W. de Vrijer, Philips tech. Rev. **36**, 305, 1976.

[16] Research on sintered-bronze bearings has been described by A. L. Braun, De Constructeur **18**, No. 5, page 61, 1979, and **18**, No. 7, page 20, 1979 (in Dutch).

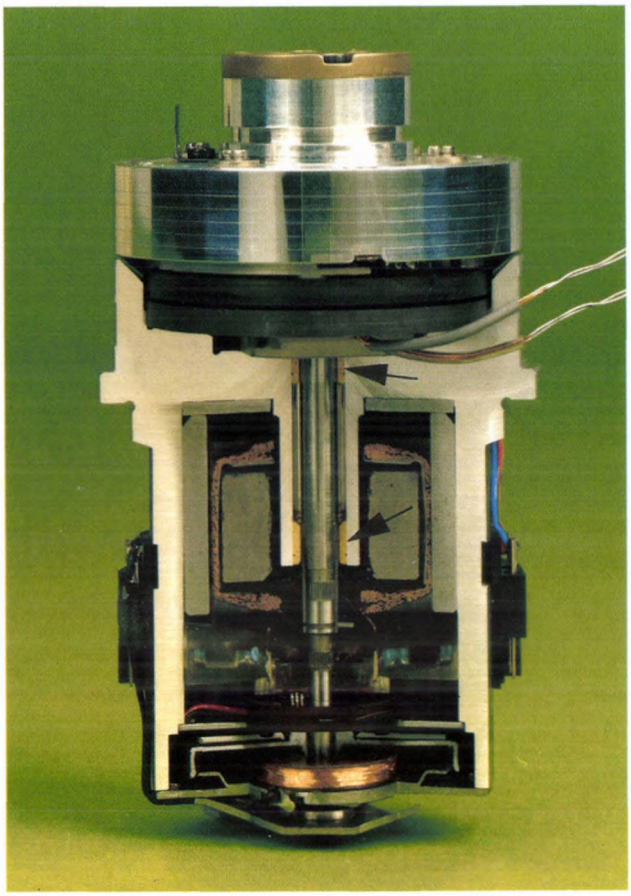


Fig. 20. Cutaway view of the VR 2020 head-disc motor. The arrows indicate the positions of the two spiral-groove journal bearings.

of the bearing bushes depend on the design of the motor. The choice of the total bearing length, the length of the groove pattern and the radial bearing gap is determined by the accuracies to which shaft and bearing bore can be made and by the requirement for sufficient stiffness at the maximum temperature (about 55 °C) and sufficiently low frictional losses at the minimum temperature (about 10 °C).

The choice of the grease is important because load-carrying capacity, stiffness, stability and frictional losses are directly proportional to the viscosity. Furthermore, the temperature dependence of the viscosity, the rheological behaviour and the life have to be adjusted to suit the specific conditions in the head-disc motor. We have obtained good results with the grease designated *ML* ('medium-light') in fig. 12, whose viscosity at 55 °C is about 9% of the viscosity at 10 °C.

Although it makes very little difference to the load-carrying capacity, the stiffness and the friction whether the grooves rotate with respect to the load vector or not, it does affect the position of the free boundary and the stability of the bearings. For manufacturing reasons we decided to make the grooves in the stationary bearing bushes. Since the tape tension provides the biggest load, this means that the grooves and the load vector do not rotate with respect to one another.

As it has already been shown, it is preferable for the section pumping upwards and the section pumping downwards to have different lengths. Furthermore the groove geometry must ensure maximum possible load-carrying capacity, stiffness and stability without the frictional losses being too high or grease leaking out of the bearings. For this to be achieved many grooves are required. *Fig. 21* is a detail from a photograph of the two bearings in the head-disc motor; it clearly shows the groove patterns.

The load-carrying capacity W and the stiffness S of the bearings are directly related. This is because the load-carrying capacity in a spiral-groove journal bearing is directly proportional to the eccentricity e and also because the angle between load vector and shaft deviation ('attitude angle') depends only on the ratio between the length and the diameter of the bearing [9]. For the stiffness we can then employ the following expression:

$$S = W/e = \frac{\eta\omega R^4 \overline{W}}{\Delta R^3},$$

where η is the viscosity of the grease, ω the angular velocity of the shaft, R the shaft radius and ΔR the radial clearance. $\overline{W}$ is a dimensionless number representing the load-carrying capacity; it is mainly dependent on the ratio of the length to the diameter of the bearing and the ratio of the groove depth to the

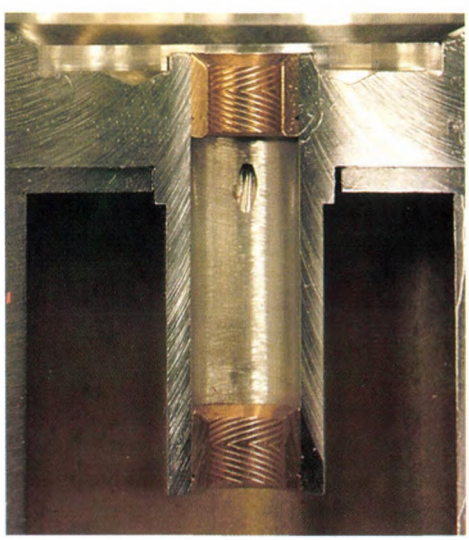


Fig. 21. Detail from photograph of two spiral-groove journal bearings in a cutaway VR 2020 head-disc motor. The grooves form an asymmetrical herringbone pattern in both bearings. Under the upper bearing can be seen the feed hole through which the grease is supplied to the lower bearing after the shaft has been mounted.

bearing clearance [9]. In *fig. 22* the calculated stiffness has been plotted against the temperature of the grease for permitted values of ΔR. The temperature dependence of ΔR has been neglected. The decrease in viscosity with increasing temperature gives a considerable reduction in the stiffness. The values of the stiffness measured for a head-disc motor with the nominal bearing gap, which have also been plotted in fig. 22, are in good agreement with the calculated values.

The frictional torque can be expressed as follows:

$$F = \frac{2\pi\eta\omega R^4 \bar{F}}{\Delta R}.$$

The dimensionless frictional torque $\bar{F}$ is proportional to the ratio of the length to the diameter, but only slightly dependent on the other groove dimensions [9]. At a given temperature the frictional torque during one revolution is constant at every instant and is practically independent of load variations. This ensures very even running for the head disc. In *fig. 23* the calculated frictional torque is shown as a function of the temperature for the permitted ΔR values. In comparison with the stiffness, the frictional torque is much less dependent on the bearing gap ($1/\Delta R$ compared with $1/\Delta R^3$) while the temperature dependence remains the same.

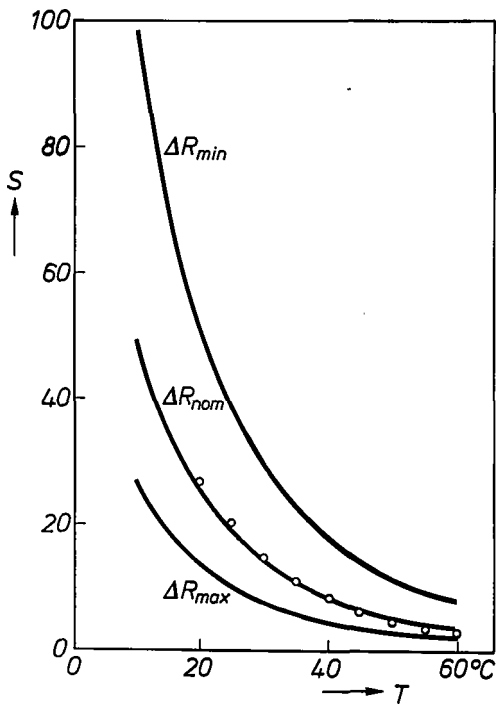


Fig. 22. Stiffness S (in arbitrary units) of spiral-groove journal bearings, calculated as a function of the temperature T of the grease, for three different values of the radial clearance. ΔR_{nom} nominal clearance; ΔR_{max} maximum clearance; ΔR_{min} minimum clearance. The values measured on a head-disc motor with nominal clearance (indicated by open circles) agree well with the calculated values for ΔR_{nom}.

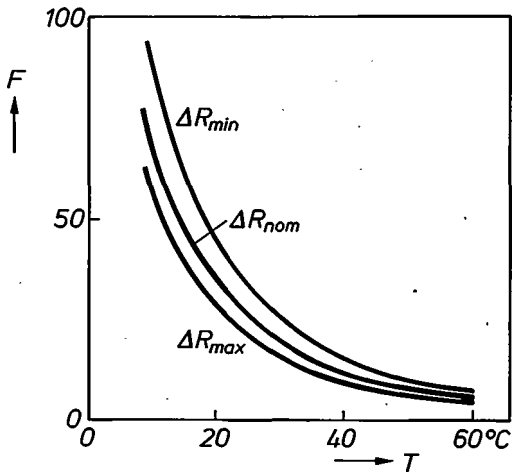


Fig. 23. Frictional torque F (in arbitrary units) of spiral-groove journal bearings, calculated as a function of the temperature T of the grease, for the three different values of the clearance ΔR. When compared with fig. 22, the effect of the clearance is much smaller but the effect of the temperature is the same.

The stability in relation to whirl depends on the radial restoring force, which ensures that the bearing returns to its equilibrium position after any disturbance. This force is at its lowest in an unloaded bearing, i.e. when the shaft is concentric with the bearing bore, so that the chance of 'half-speed whirl' is highest. Unlike plain journal bearings, spiral-groove journal bearings can also be stable in an unloaded state. A bearing will be stable if the condition

$$\frac{M\omega\Delta R^3}{\eta R^4} < \bar{M}_c$$

is satisfied, where M is the part of the rotor mass that affects the stability of the bearing and $\bar{M}_c$ the dimensionless number representing the stability of the bearing. *Fig. 24* illustrates the movement of the shaft that was observed after starting an unloaded head-disc motor. The shaft, which rests against an arbitrary point on the bearing wall when the motor is stationary, finds its equilibrium position very soon.

Finally, we should make a few comments about the life of the spiral-groove journal bearings in the head-disc motor. Because there is metal contact between the shaft and the bearing wall in these bearings only when they are at rest, the life of the bearings is mainly determined by the life of the grease. As we saw in our discussion of the greases, the requirements relating to the viscosity, flow behaviour and oxidation resistance are the important ones. In the grease we used, these properties do not change much with operating time,

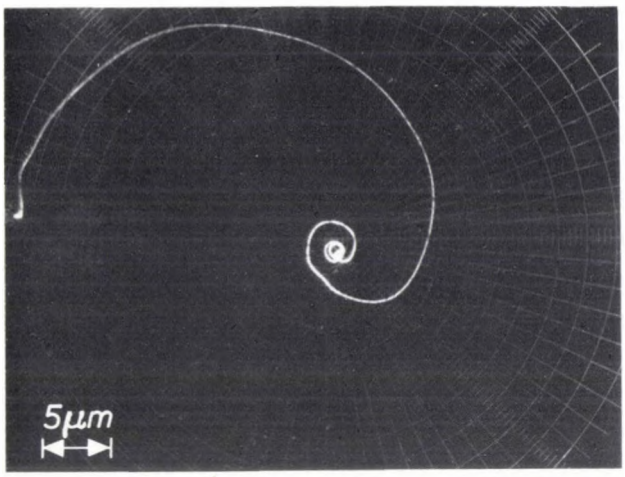


Fig. 24. Photograph illustrating the displacement of the shaft after an unloaded head-disc motor has been started. The equilibrium position in which shaft and bearing bush are concentric is reached very soon.

and so the bearings can have a very long life. In life tests of 3000 hours of operation simulating 'consumer' conditions we have observed no deterioration at all in the bearing behaviour.

Summary. The operation of spiral-groove bearings depends on the groove dimensions, the groove pattern and on whether the grooves rotate with respect to the load vector or not. The lubricant is also important: bearing characteristics such as load-carrying capacity, stiffness and frictional torque are proportional to the viscosity. If a spiral-groove bearing is lubricated with grease it can have a high load-carrying capacity and even at rest there will be practically no leakage of lubricant. The special conditions that occur in a spiral-groove bearing often place exacting requirements on the grease, particularly for its rheological behaviour and its stability during long periods of use. Spiral-groove bearings can be made from both metal and plastic. Different methods can be used for making the bearing bore and forming the grooves. Because of their special features grease-lubricated spiral-groove journal bearings have recently been incorporated in a consumer article: they form the journal bearings for the head-disc motor in the VR 2020 video cassette recorder, based on the new 'Video 2000' system.

Scientific publications

These publications are contributed by staff of laboratories and plants which form part of or cooperate with enterprises of the Philips group of companies, particularly by staff of the following research laboratories:

Philips Research Laboratories, Eindhoven, The Netherlands	*E*
Philips Research Laboratories, Redhill, Surrey RH1 5HA, England	*R*
Laboratoires d'Electronique et de Physique Appliquée, 3 avenue Descartes, 94450 Limeil-Brévannes, France	*L*
Philips GmbH Forschungslaboratorium Aachen, Weißhausstraße, 51 Aachen, Germany	*A*
Philips GmbH Forschungslaboratorium Hamburg, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany	*H*
Philips Research Laboratory Brussels, 2 avenue Van Becelaere, 1170 Brussels (Boitsfort), Belgium	*B*
Philips Laboratories, N.A.P.C., 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A.	*N*

B. Aldefeld: A numerical solution of transient nonlinear eddy-current problems including moving iron parts.
IEEE Trans. **MAG-14**, 371-373, 1978 (No. 5). *H*

S. Askienazy (Hôpital Ste Anne, Paris), **R. Genève & M. Jatteau:** Les tendances de la médecine nucléaire.
Acta Electronica **22**, 79-90, 1979 (No. 2). *L*

M. Auphan: Les transducteurs à réseau annulaire focalisant en poursuite d'échos.
Acta Electronica **22**, 119-127, 1979 (No. 2). *L*

C. I. M. Beenakker, B. Bosman & P. W. J. M. Boumans: An assessment of a microwave-induced plasma generated in argon with a cylindrical TM_{010} cavity as an excitation source for emission spectrometric analysis of solutions.
Spectrochim. Acta **33B**, 373-381, 1978 (No. 7). *E*

A. J. van Bommel & J. E. Crombeen: The GaP(001) surface and the adsorption of Cs.
Surface Sci. **76**, 499-508, 1978 (No. 2). *E*

A. H. Boonstra & C. A. H. A. Mutsaers: Small values of the temperature coefficient of resistance in lead rhodate thick films ascribed to a compensation mechanism.
Thin Solid Films **51**, 287-296, 1978 (No. 3). *E*

G. Bouwhuis & J. J. M. Braat: Video disk player optics.
Appl. Optics **17**, 1993-2000, 1978 (No. 13). *E*

J. J. M. Braat & G. Bouwhuis: Position sensing in video disk readout.
Appl. Optics **17**, 2013-2021, 1978 (No. 13). *E*

J. J. M. Braat & G. Bouwhuis: Optical video disks with undulating tracks.
Appl. Optics **17**, 2022-2028, 1978 (No. 13). *E*

J. C. Brice: Crystal growth from liquids at high temperatures.
Progress in Crystal Growth and Characterization **1**, 255-288, 1978 (No. 3). *R*

J. W. Broer: Linking while writing — do you or don't you?
J. tech. Writ. Comm. **8**, 217-225, 1978 (No. 3). *E*

R. Bruno, W. Hermann, H. Hörster, R. Kersten, K. Klinkenberg & F. Mahdjuri: The Philips experimental house: results and experience.
Energy use management, Proc. int. Conf., Tucson 1977, pp. 299-304. *A*

R. Bruno & R. Kersten: Models and methods for the analysis and optimization of solar energy systems.
Energy use management, Proc. int. Conf., Tucson 1977, pp. 643-651. *A*

A. L. J. Burgmans & M. F. H. Schuurmans: Spectroscopie aan atomen nabij een wand.
Ned. T. Natuurk. A **44**, 62-64, 1978 (No. 2). *E*

T. A. C. M. Claasen: Warum Digitaltechnik bei der Übertragung und Speicherung von Bild- und Tonsignalen?
Fernseh- u. Kino-T. **32**, 245-250, 1978 (No. 7). *E*

P. S. Clarke, J. W. Orton & A. J. Guest: Electrical-conductivity and Hall-effect measurements in semiconducting powders. Study of percolation effects.
Phys. Rev. B **18**, 1813-1817, 1978 (No. 4). *R*

R. H. Coursant: Les transducteurs ultrasonores.
Acta Electronica **22**, 129-141, 1979 (No. 2). *L*

P.-J. Courtois: Exact aggregation in queueing networks.
First Meeting AFCET-SMF on Applied Math., Palaiseau 1978, Vol. I, pp. 35-51. *B*

P. A. Devijver: A generalized divergence measure for signal selection.
First Meeting AFCET-SMF on Applied Math., Palaiseau 1978, Vol. I, pp. 315-324. *B*

H. Dötsch: Investigation of magnetic bubble films using spin wave and magnetoelastic resonances.
IEEE Trans. **MAG-14**, 692-694, 1978 (No. 5). *H*

A. Dumont: Les techniques Doppler en échographie.
Acta Electronica **22**, 143-148, 1979 (No. 2). *L*

R. C. French: Error rate predictions and measurements in the mobile radio data channel.
IEEE Trans. **VT-27**, 110-116, 1978 (No. 3). *R*

R. Genève: Introduction to medical imaging techniques.
Acta Electronica **22**, 7-17, 1979 (No. 1). *(Also in French.)* *L*

J. P. Hazan, J. P. Cabanié & J. J. Bernard: Method of assessing index profile data.
Electronics Letters **14**, 416-418, 1978 (No. 14). *L*

J. P. J. Heemskerk: Noise in a video disk system: experiments with an (AlGa)As laser.
Appl. Optics **17**, 2007-2012, 1978 (No. 13). *E*

B. A. J. Jacobs: Laser beam recording of video master disks.
Appl. Optics **17**, 2001-2006, 1978 (No. 13). *E*

M. Jatteau, P. Lelong, G. Normand, J. Ott, J. Pauvert & J. Pergrale: Pour une optimisation des caméras de gammagraphie de type Anger.
Acta Electronica **22**, 91-117, 1979 (No. 2). *L*

W. H. de Jeu & F. Leenhouts (University of Groningen): Physical properties of nematic p,p'-diheptylazobenzene.
J. Physique **39**, 869-872, 1978 (No. 8). *E*

H.-G. Junginger, R. F. Schmidt & R. Strunk: Experiments on latent electrostatic images and electrophoretic development comparison with model calculations.
Photogr. Sci. Engng. **22**, 213-218, 1978 (No. 4). *A*

J. J. Kelly & G. J. Koel: Electrochemical aspects of the beveling of sputtered Permalloy films.
J. Electrochem. Soc. **125**, 860-865, 1978 (No. 6). *E*

G. Kowalski, R. Rieckeheer & W. Wagner: New means for picture formation in computer tomography.
Acta Electronica **22**, 51-63, 1979 (No. 1). *H*

C. Kramer (Philips Medical Systems Division, Best): Developments in medical imaging techniques.
Acta Electronica **22**, 19-39, 1979 (No. 1). *(Also in French.)*

S. Leblanc, M. Bristica (La Radio-technique, Suresnes) **& V. Evers:** Experience with an undergraduate level CAI-course for electronic engineers in industry.
Computers & Education **2**, 221-226, 1978 (No. 3). *L, E*

K.-M. Luedeke, J. Koehler & J. Kanzenbach: A radiation balance microwave thermograph for medical applications.
Acta Electronica **22**, 65-69, 1979 (No. 1). *H*

P. J. Mabey: Mobile radio data transmission — coding for error control.
IEEE Trans. **VT-27**, 99-109, 1978 (No. 3). *R*

A. E. Morgan & H. W. Werner: Semiquantitative analyses by secondary ion mass spectrometry using one fitting parameter.
Mikrochim. Acta 1978 II, 31-50 (No. 1/2). *E*

A. Nicia: Practical low-loss lens connector for optical fibres.
Electronics Letters **14**, 511-512, 1978 (No. 16). *E*

A. Pirotte: Criteria for comparing the linguistic structure of non-procedural relational languages.
Proc. Workshop 'Modèles relationnels — Langages de manipulation de données', Paris 1978, pp. 228-249. *B*

J. Polman, A. K. de Jonge & A. Castelijns: Free piston electrodynamic gas compressor.
Proc. 1978 Purdue Compressor Technology Conf., West Lafayette (Indiana), pp. 241-245. *E*

H. Rau: Nonstoichiometry of ZnSe and CdSe.
J. Phys. Chem. Solids **39**, 879-882, 1978 (No. 8). *A*

N. V. Smith (Bell Laboratories, Murray Hill, N.J.) **& P. K. Larsen:** Directional photoemission from two-dimensional systems.
Photoemission and the electronic properties of surfaces, ed. B. Feuerbacher, B. Fitton & R. F. Willis, pp. 409-436; Wiley, New York 1978. *E*

W. T. Stacy & B. J. Fitzpatrick: Electron-beam-induced dislocation climb in ZnSe.
J. appl. Phys. **49**, 4765-4769, 1978 (No. 9). *E, N*

T. Thalhammer, G. Prast, J. Langerhorst & W. Floor (Ministry of Foreign Affairs, The Hague): Development co-operation and solar energy.
Int. J. Energy Res. **2**, 211-228, 1978 (No. 3). *E*

J. C. Tranchart, L. Hollan & R. Memming: Localized avalanche breakdown on GaAs electrodes in aqueous electrolytes.
J. Electrochem. Soc. **125**, 1185-1187, 1978 (No. 7). *L*

C. H. F. Velzel: Laser beam reading of video records.
Appl. Optics **17**, 2029-2036, 1978 (No. 13). *E*

J. Verhoeven: A vapor collect method applied to the measurement of magnesium diffusion in tungstened nickel.
Appl. Phys. Letters **33**, 16-18, 1978 (No. 1). *E*

J. H. Waszink & H. J. Flinsenberg: Determination of the electron density in a high-pressure Na-Xe discharge from the profile of a Stark-broadened spectral line.
J. appl. Phys. **49**, 3792-3795, 1978 (No. 7). *E*

H. Weiss, E. Klotz & R. Linde: Flashing tomosynthesis: three-dimensional X-ray imaging.
Acta Electronica **22**, 41-50, 1979 (No. 1). *H*

J. Wilson, D. C. Brown (both with University of Rochester, N.Y.) **& W. K. Zwicker:** XeF excimer pumping of $Nd:P_5O_{14}$.
Appl. Phys. Letters **33**, 614-616, 1978 (No. 7). *N*

J. P. Woerdman: Laser-excited broadband violet emission from sodium molecules.
Optics Comm. **26**, 216-218, 1978 (No. 2). *E*

C. E. C. Wood & B. A. Joyce: Tin-doping effects in GaAs films grown by molecular beam epitaxy.
J. appl. Phys. **49**, 4854-4861, 1978 (No. 9). *R*

Supplement to Philips Technical Review
Beilage der Philips Technischen Rundschau
Bijlage van Philips Technisch Tijdschrift

PHILIPS

February 1981

Recent United States Patents

Abstracts from patents that describe inventions from the following research laboratories that form part of or cooperate with the Philips group of companies:

Philips Research Laboratories, Eindhoven, The Netherlands	*E*
Philips Research Laboratories, Redhill, Surrey RH1 5HA, England	*R*
Laboratoires d'Electronique et de Physique Appliquée, 3 avenue Descartes, 94450 Limeil-Brévannes, France	*L*
Philips GmbH Forschungslaboratorium Aachen, Weißhausstraße, 51 Aachen, Germany	*A*
Philips GmbH Forschungslaboratorium Hamburg, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany	*H*
Philips Research Laboratory Brussels, 2 avenue Van Becelaere, 1170 Brussels (Boitsfort), Belgium	*B*
Philips Laboratories, N.A.P.C., 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, USA	*N*

4 207 595

Apparatus for making laminar radiograms

J. Dittrich *H*
J. Heinzerling
P. Lux

In tomosynthesis distortions are produced owing to the curvature of the image intensifier input screen and other non-linearities of the transmission system, which in particular affect the synthesis of the laminar image when the individual images should be shifted relative to each other for this purpose. These distortions depend on the angle which the central ray makes with the laminar plane during the individual exposures and therefore they can hardly be compensated for. The invention now proposes that when the individual images are taken the image intensifier is steered so that the optical axis always remains parallel to the central ray. The distortions produced by the curvature of the input screen of the image intensifier etc. are then independent of the angle between the central ray and the laminar plane, and can be compensated for by known commercially available correction devices. However, additional geometric distortion is then produced because the plane of the input screen is inclined relative to the laminar plane. These geometric distortions depend on the orbit angle. By a rotation of the deflection field in accordance with the orbit angle, it is then achieved that trapezium distortions are produced which are independent of the orbit angle. These trapezium distortions are compensated for in known manner, and subsequently the position of the deflection field is restored.

---

4 207 614

Magnetic bubble shift register store

G. Frens *E*

A magnetic bubble shift register store having a plate of magnetic material whose preferred magnetization direction extends transverse to the plane of the plate and in which bubbles are situated, said plate having two separate, elongated generally parallel extending continuous bubble paths, of lower bubble energy in comparison with the vicinity; the center lines of said paths being situated at a distance from each other which is at least equal to the mean bubble diameter. The 0-bits of written information to be transported and stored are represented by bubbles in the one path, while the 1-bits thereof are represented by bubbles in the other path. The interaction between the bubbles ensures that bubbles in the two paths cannot pass each other, with the result that the information represented by the bubbles can be unambiguously transported and maintained in the path direction.

4 207 656

Color television display tube and method of manufacturing same

J. van Esdonk *E*
P. F. A. Haans

In an assembly of at least two electrodes which are connected together in an insulating manner, said electrodes are kept at a given distance from each other by at least one member of an electrically insulating material situated between the electrodes which comprises a core which determines the distance between the electrodes and a jacket which directly adheres to the electrode material by heating. The material of the core has a higher melting point than the material of the jacket so that the core during effecting a connection between the jacket and the electrode material maintains its shape.

---

4 208 449

Method of making an electric resistor having a resistance body consisting of silicon carbide having a negative temperature coefficient

W. F. Knippenberg *E*
G. Verspui
S. H. Hagen

A method of making an electric resistor having a negative temperature coefficient of resistance whose resistance body consists of *P*-type doped pyrolytic polycrystalline cubic silicon carbide.

---

4 208 607

Electric lamp with tin or lead alloy plug for lead-in

P. Hokkeling *E*
R. J. Q van den Plas

The invention relates to an electric lamp having a pinch seal in which a molybdenum foil is incorporated as a current leadthrough conductor. An external current conductor is connected to the foil. As a result of differences in coefficients of expansion, a capillary space is present around the external current conductor, through which space oxidizing gas can reach the molybdenum foil. Oxidation of the foil involves an increase of its volume and results in crack of the pinch seal. In lamps according to the invention a metal

plug is provided around the external current conductor, said plug being sealed to said current conductor and to the glass of the pinch seal and sealing the capillary space in a vacuum-tight manner.

---

4 208 634

Circuit for suppressing noise caused by scratches on a phonograph record

J. B. H. Peek E
J. M. Schmidt

A circuit for suppressing pulse-shaped interferences in an audio signal caused by scratches on a phonograph disc, comprising a signal processing section connected between an audio frequency input and an audio frequency output, comprising a noise suppressor having a control input; as well as a control signal section connected between the audio frequency input and the control input, said control signal section comprising a first threshold selection circuit connected to the audio frequency input for selecting pulse-shaped signals from the audio signal, followed by a second threshold selection circuit for distinguishing the pulse-shaped signals originating from noise signals from those originating from musical signals.

---

4 209 804

Record carrier containing information in an optically readable radiation reflecting information structure

J. G. Dil E

A record carrier is described having an optically readable radiation-reflecting information structure, comprising information areas arranged in information tracks, which areas are spaced from each other by intermediate areas, the information areas having oblique walls. It is demonstrated that a suitable information signal and a suitable positional error signal are obtained if the angle of inclination of the walls of the information area lies between 65° and 85° and the phase depth of the information areas lies between 95° and 140°.

---

4 210 371

Rotary-anode X-ray tube

J. Gerkema E
E. A. Muijderman

A rotary-anode X-ray tube, comprising a rotary anode which is journalled in a vacuum-tight housing by means of at least one sleeve bearing whose mutually co-operating bearing surfaces are made of W or Mo or of an alloy of W and Mo. The bearing is lubricated by Ga or a Ga alloy having a melting point below 25 °C. The Ga or Ga alloy lubricant is in molecular wetting contact with the bearing surfaces so that the bearing surfaces are completely separated from each other in the loaded condition. During standstill, as well as during rotation, the lubricant is not forced out of the bearing, so that the bearing surfaces cannot become fused. The wear and the bearing noise are therefore extremely low. The desired, thorough wetting is obtained by making the bearing surfaces, as well as the Ga alloy, oxide-free before they are brought into contact with each other.

---

4 210 882

Delay network comprising a chain of all-pass sections

E. Roza E
J. O. Voorman

A delay network, having a chain of all-pass sections, each comprising two separate branches, a resistive and a capacitive branch, which terminate in amplifiers with negligible signal consumption whose output signals are combined. This enables an analog delay network to be realized in integrated circuit technology.

4 213 100

Gas discharge laser

R. A. J. Keijser E
G. A. Wesselink
B. J. Derksema
J. A. T. Verhoeven

A gas discharge laser in which at least the inner wall of the laser tube is manufactured for the greater part from a gehlenite glass thereby providing a laser having a long life. The gehlenite glass can better withstand the eroding effect of the gas discharge so that the rate at which the optical elements are contaminated is greatly reduced.

---

4 213 144

Method of modulating a composite color television signal on a carrier signal and device for carrying out said method

E. de Boer E

A method of modulating a composite color television signal on a carrier wave. The luminance signal is frequency-modulated on the carrier wave. The chrominance signal is translated to a subcarrier wave frequency either by mixing the modulated carrier wave or frequency modulating the carrier wave with said chrominance signal. Subsequently, the carrier wave is pulse-width modulated with said subcarrier wave. In this way a signal is obtained from which the original color television signal can be recovered by a single frequency demodulation operation, while the normally occurring undesired interference components can be reduced substantially.

---

4 213 311

Superleak

A. P. Severijns E
F. A. Staas

A superleak in which a heat exchanger having a superleak structure is included in order to drastically reduce heat leak in the flow direction during operation.

---

4 213 659

Connecting rod bearing arrangement

L. P. M. Tielemans E
J. A. G. de Deugd

A bearing arrangement for a connecting rod big end journalled on a crankshaft pin, for lubrication by splash or mist. Adjoining the cylindrical journal on the crankshaft pin and the rod big end, facing generally planar surfaces on the crankshaft and the rod big end have a pattern of shallow lubricant pumping grooves to provide a constant, reliable flow of oil through the bearing. At the opposite end of the connecting rod plain bearing, other plane surfaces may provide either draining grooves or grooves which pump oil outward to an axial bearing surface.

---

4 216 004

Method of breaking optical fibers

R. Brehm E
A. J. J. Franken

A method of breaking optical fibers, a fiber being scored over the entire circumference in a plane perpendicular to the fiber axis. By subsequently applying a predetermined axial tensile force to the fiber, fracture is initiated over the entire circumference of the fiber. As a result of this, fibers with a comparatively large diameter can be broken in such a way that a mirror zone is obtained across the entire fracture area. Circumferential scoring also ensures the required accuracy in respect of the perpendicular orientation of the fracture plane relative to the fiber axis.

4 216 274

Battery with hydrogen absorbing material of the formula L_nM_5

H. A. C. M. Bruning E
J. H. N. van Vucht
F. F. Westendorp
H. Zijlstra

An electrochemical energy conversion device, i.e. a galvanic cell using a hydrogen absorbing material to provide either a source of energy, or to absorb hydrogen to avoid overpressure in the cell. This hydrogen absorbing material is a compound of the formula L_nM_5 in which L_n represents a lanthanide metal and M is either cobalt or nickel.

---

4 216 419

Tachometer system

R. A. A. F. van Dam E
K. A. Immink

A tachometer system for supplying a control signal indicative of positional and/or speed errors of a rotatable element comprises a tachometer on which marks are arranged in a closed track and a detector responsive thereto to produce n tacho-pulses per revolution of the tachometer. The tachometer system further comprises a memory device with n locations in which n correction signals are stored. The correction signals are obtained by driving the tachometer at an accurately constant speed, measuring the phase difference between the tacho-pulses and the reference signal, and storing the phase-error signals related to the consecutive tacho-pulses in separate location of the memory device. When the tachometer system is used in a servo control loop these correction signals are used to correct the phase-error signals then measured, yielding a correction in respect of the deviations in these phase-error signals owing to positional tolerances of the marks on the tachometer.

---

4 217 489

Device for location-sensitive detection of photon and/or particle radiation

J.-C. Rosier L

An electron multiplier which is formed by at least two microchannel plates which are arranged one above the other and which operate in the saturation mode. The second microchannel plate produces an output charge for each incident electron which depends on the electrical field in the vicinity of the input surface of said second plate. By suitable geometrical structure in the plate or by shaping the potential applied thereto, an electrical field is produced which varies from point to point on its input surface. The charge output thus contains information regarding the location of the incident electron.

---

4 218 269

Method for the epitaxial deposition of several layers

T. G. J. van Oirschot E
W. J. Leswin
P. J. A. Thijs
W. Nijman

A method of manufacturing a semiconductor device having a monocrystalline substrate and a plurality of epitaxial layers successively deposited on the substrate is disclosed. The device is manufactured by successively contacting the substrate with solutions which are previously saturated by contact with a plurality of auxiliary substrates, in a process in which the monocrystalline substrate, the auxiliary substrates and the solutions are cooled before the layers are deposited. The method includes the steps of contacting a first auxiliary substrate with a first solution, contacting a second auxiliary substrate with this first solution while simultaneously contacting the first auxiliary substrate with a second solution, contacting the monocrystalline substrate with the first solution to deposit a first layer thereon while simultaneously contacting a second auxiliary substrate with the second solution, and then contacting the monocrystalline substrate with the second solution to deposit a second layer thereon. The total contact time of each solution with the auxiliary substrates is not limited to the time of a single contact of the monocrystalline substrate with any solution. In this manner, the solubility times for the auxiliary substrates need not be limited to the epitaxial layer growth time of the monocrystalline substrate, thus resulting in an improved and more flexible manufacturing process.

---

4 218 762

Device for close-packed magnetic domains

J. Haisma E
K. L. L. van Mierloo

A device for storing digital information in the form of magnetic domains including at least two ferrimagnetic layers on a substrate. The ferrimagnetic layers are separated by a compensation wall and in the presence of a bias magnetic field single magnetic domains and superposed magnetic domain pairs, the superposed domains being separated by a compensation wall, are generated. All single and superposed domains are mutually repulsive. The domains within a superposed pair are attracted to each other. The generated domains are propagated to a shift register and when the register is filled a row of domains is coincidentally propagated, transversely out of the register, to a domain array region adjacent the register. Rows are then propagated through the array region to a second shift register where they are individually detected and annihilated. A control device assures the proper working relationship between all elements.

---

4 219 321

Device for making discs

P. L. Holster E
A. Ooms

A device for making discs, comprising a press which includes at least one hydraulic pressure chamber, one wall of which is formed during operation by a flexible, thermally conductive plate which can be pressed against a thermoplastic substrate. The device comprises a first system of ducts which includes a pump whereby a hot liquid can be supplied to and discharged from the pressure chamber, and a second system of ducts which includes a pump whereby cold liquid can be supplied to the pressure chamber. The liquid in the systems of ducts and the pressure chamber is pressurized by means of an expansion vessel, the liquid part of which communicates with the pressure chamber as well as with both systems of ducts, its gas part communicating with a pressure control device.

---

4 220 844

Method of and device for plasma-MIG welding

W. G. Essers E

A plasma-MIG welding system includes establishing a plasma arc between a primary non-consumable electrode and an auxiliary non-consumable nozzle electrode to provide a plasma flow downstream of the primary non-consumable electrode and out through such nozzle electrode. A consumable electrode is fed through the plasma flow out through the nozzle electrode, a MIG-arc being established between the consumable electrode and a workpiece. A gas stream consisting of an inert gas and an oxidizing gas is flowed past the consumable electrode in a manner to completely envelope the same and into admixture with the plasma flow without coming into contact with the non-consumable electrode. A separate gas stream provides a sheath for the resulting plasma flow after its passage through the nozzle electrode.

---

4 221 928

Noise reduction circuit for stereo signals

N. V. Franssen E
M. H. Geelen

Noise reduction circuit for stereo signals comprising a control input terminal, first and second outputs as well as a controllable coupling circuit for a controllable mutual signal coupling between the two outputs, the two outputs being decoupled from one another only during those sound passages wherein an effective after-effect of a stereo impression is produced to obtain an optimal combination of noise level and stereo effect.

4 221 988

Low pressure gas discharge lamp having fibers evenly distributed between the electrodes

J. Hasker E
P. R. van IJzendoorn
H. Roelofs

A low pressure gas discharge lamp having a body present in the discharge vessel which consists of a longitudinal support which extends into the longitudinal direction of the vessel, the support being provided with fibers which are distributed over the space within the discharge vessel and extend substantially transversely from the support.

---

4 222 102

Data buffer memory of the 'first-in, first-out' type, comprising a variable input and a variable output

P. G. Jansen E
J. L. W. Kessels
B. L. A. Waumans

A data buffer memory of the 'first-in, first-out' type, having an input bus by which data are applied to the buffer and an output bus by which data are taken up from the buffer. The buffer includes logic means whereby a variable input location and a variable output location can be selected. The logic means are provided for each section of the buffer and use status signals in co-operation with signals applied from outside the buffer, to determine where data are to be written in the buffer and where data are to be read from the buffer.

---

4 222 159

Method of manufacturing a color display tube shadow mask

J. Koorneef E

A method of manufacturing a color display tube of the refocusing type in which supports of insulation material are secured against an apertured metal plate. The supports are provided with a conductor at least on the side remote from the plate, so that the plate constitutes a first set of lens electrodes and the conductors constitute a second set of lens electrodes. The two sets of lens electrodes form a quadrupole lens in each aperture in the metal plate when a voltage difference is applied between the first set and the second set. The defocusing direction of the quadrupole lens is parallel to the phosphor strips of the display screen.

---

4 222 791

Method of making heterojunction devices by accurate control of gold pattern mask and proton bombardment

R. P. Tijburg E
T. van Dongen

A method of manufacturing a device in which a surface of a body is provided with a gold layer masking against a proton bombardment, in which the gold pattern is formed by means of a photoetching process, the gold pattern is removed after the subsequent proton bombardment of the body. The masking gold layer is obtained by providing on the body a quantity of gold simultaneously with a quantity of an addition which is small as compared with the quantity of gold.

---

4 222 827

Electroplating solution for the electrodeposition of aluminium

T. E. G. Daenen E

Electrolyte liquid for electrodepositing ductile aluminium. The liquid comprises, dissolved in an aprotic liquid, an alkali aluminium hydride and aluminum hydride coordinatively bound to a tertiary amine, a tertiary diamine or an arylphosphine.

4 223 282

Method of reducing interference components in a frequency modulated signal and device for carrying out said method

E. de Boer E

A method of reducing interference components in a frequency-modulated carrier signal in which the modulating signal contains a composite color television signal with a chrominance signal modulated on a chrominance subcarrier wave. In addition to the frequency modulation, the carrier signal is pulse-width modulated by the first-order lower side-band component of the frequency-modulated signal produced by the modulated chrominance subcarrier signal.

---

4 223 347

Videodisc with undulating nested tracks

G. Bouwhuis E
P. Kramer

A record carrier is described on which information is stored in an optically readable structure of trackwise arranged areas and intermediate areas. By giving the information tracks periodic excursions towards and away from adjacent tracks, in a direction transverse to the direction of reading, the period of the excursions being substantially greater than the average period of the areas and the amplitude of the excursions being smaller than the width of the tracks, a control signal for centering the read beam of radiation on an information track can be obtained without the use of additional optical means. The excursions of the tracks have such relative phase that the tracks are in nested concentric relation and one component of the stored information represents the phase of the undulations.

---

4 225 885

Method and apparatus for adaptive transform coding of picture signals

P. Lux H
J. Petzold

A method and apparatus to reduce the redundancy for the transmission or storage of picture signals using transforms. The basis coefficients are arranged in a sequence and thereafter quantized in the customary manner. The basis coefficients obtained from the transformation are sequentially examined to see if they exceed a predetermined value, for example the average value across the entire picture. The number of times the value is exceeded is counted. If, in a sequence class, the number of coefficients exceeding the predetermined value is more than half the total number in the sequence class then the associated sequences are quantized with a larger number of bits. For this purpose, activity classes are formed which indicate which sequences in the picture are active and the characteristic of the quantizer is controlled by this class information.

---

4 226 227

Solar collector

G. Harupa A

There is provided a solar collector including a sealed and evacuated transparent cover tube that forms a wall of a heating duct through which a heat transport medium is fed during operation. Special connections are provided at the ends of the heating duct for the introduction of the heat transport medium thereinto and the discharge of the same therefrom.

---

4 226 509

Display device comprising a liquid display medium

J. H. Jacobs E

In a display device having a liquid display medium, for example liquid crystal, the filling apertures have to be readily sealed. In a seal which is very suitable for mass production a piece of elastic material is drawn through two oppositely located filling apertures or between a brace across the filling aperture and the filling aperture.

PHILIPS TECHNICAL REVIEW

VOLUME 39, 1980, No. 8

A new concept for television camera tubes

J. H. T. van Roosmalen

An investigation into the possibilities of making small colour-television cameras has resulted in a new concept for camera tubes. The new tube has functional and simple electron optics and can be produced with great precision by means of a new technology. On the basis of this concept very small experimental camera tubes have been made, which readily meet the exacting requirements for professional applications.

Operation of a camera tube

In a television camera tube an electron gun scans a photoconducting layer that carries on the other side an image of a scene [1]; see *fig. 1*. Between this layer and the glass window there is a transparent conducting layer, the signal plate. In the unexposed state the potential of the free surface of the photoconducting layer is approximately equal to the potential of the cathode of the electron gun. A light pattern from the scene produces a potential pattern (from 0 to say 5 V) on this surface, since charge is drawn away to the signal plate through the layer in correspondence with the light intensity at each location. The scanning electron beam replenishes the charge again and produces the television signal in the form of a voltage across the output resistor.

To reproduce colours faithfully a professional camera for colour television has *three* camera tubes, for the primary colours red, green and blue.

The electron beam that scans the photoconducting layer is first focused and deflected. In our experimental tube the focusing is electrostatic and the deflection is magnetic. We chose such an arrangement largely because of the advantages such as light weight and small size, as mentioned earlier in this journal [2,3].

To permit a faithful reproduction of rapidly changing scenes, a camera tube should have a rapid response (i.e. minimum lag). This is partly determined by the time required to charge the photoconducting layer, and decreases with the capacitance of the layer and the resistance of the scanning electron beam. The capacitance is inversely proportional to the thickness of the layer. However, depending on the type of layer, disadvantages of a thicker layer are that the image produced is less sharp because of the increased diffu-

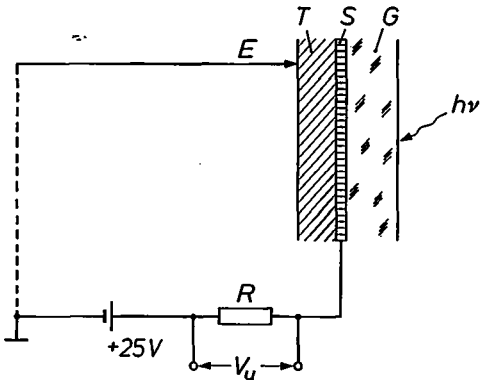


Fig. 1. Schematic operation of a television camera tube. The electron beam E scans the photoconducting layer (the target) T (e.g. in a Plumbicon[*] tube). A light pattern from the scene falls on the other side, where it produces a potential pattern on the free surface of T because charge leaks away to the conducting transparent signal plate S (typically of SnO_2), which has a potential of say 25 V. The scanning beam replenishes the charge deficit, so that a voltage V_u, the television signal, appears across the resistor R. G glass window.

J. H. T. van Roosmalen is with Philips Research Laboratories, Eindhoven.

[*] Registered Trade Mark for television camera tubes.

[1] A detailed description is given in E. F. de Haan, A. van der Drift and P. P. M. Schampers, Philips tech. Rev. **25**, 133, 1963/64.

[2] J. H. T. van Roosmalen, Philips tech. Rev. **28**, 60, 1967.

[3] P. H. Broerse, J. H. T. van Roosmalen and S. L. Tan, Philips tech. Rev. **29**, 325, 1968.

sion of the light, or that the photoconductive lag caused by charge trapping increases, so that the choice of layer thickness entails a compromise. The beam resistance R_b is given by

$$R_b = \frac{kT_b}{eI_a}, \quad (1)$$

where k is Boltzmann's constant, e is the electronic charge and I_a is the part of the beam current accepted by the layer. The beam temperature T_b is determined by the axial velocity distribution of the scanning electrons.

A problem that arises in conventional camera tubes is that the beam temperature is too high, and this introduces a limitation to the speed of response. The high beam temperature is due to a strong interaction between the electrons, which appears to be closely related to the use of a triode as electron gun.

New light on old problems

A few years ago in our laboratories we started looking for a fundamental solution to the problems encountered with camera tubes, with the object of making it possible to produce better and smaller tubes. The investigation led to a new electron-optical design, in which the most important step was the replacement of the triode by a diode [4]. In the triode used in conventional camera tubes the anode has a relatively high potential (usually about 300 V), whereas the anode potential of the diode is very much lower: about 30 V. As we shall show, the lower anode potential and the lower maximum current density in a diode substantially reduce the number of electrons that are 'too fast', thereby greatly increasing the speed of response. The electron beam from a diode with a low anode potential is rather divergent and therefore cannot easily be combined with a conventional electrostatic focusing lens. A second major step was therefore the use of an accelerating lens for the electrostatic focusing. In this scheme the anode of the diode forms a single unit with the low-potential electrode of this lens. With this accelerating lens, combined with a strong landing-correction lens, very sharp images are obtained. Added advantages are that residual gases in the tube have little effect on the cathode emission, and that no provisions are necessary for an alignment of the electron beam.

For electrostatic focusing a number of electrodes insulated from one another are required, whose shape and position must be well defined. The use of conventional tubular electrodes, mechanically connected by means of insulating rods, makes it very difficult however to line up the electrodes and keep them in an aligned position, especially during temperature changes. We have therefore developed a technology in which the electrodes consist partly of a thin metal film on the inside wall of an accurately formed glass envelope. The remaining electrode assembly is carried out very accurately on locating surfaces on the inside of the envelope.

The combination of the new electron-optical design and the new technology makes it possible to produce compact camera tubes that have excellent characteristics. Before dealing with these, we shall look first at the electron optics of the new design and briefly describe the technology.

A diode as electron gun

The axial velocity distribution of the electrons from an electron gun can be determined by measuring the 'beam acceptance' curve for normal incidence of the electron beam at a conducting plate. This beam acceptance is defined as the current accepted by the target as a function of its potential with respect to the cathode. The velocity distribution of the electrons leaving the central part of the cathode in the axial direction approximates to a Maxwellian velocity distribution; it depends on the temperature T_c of the cathode [5]. With such a distribution the accepted current I_a is given as a function of the target potential V_t by

$$I_a = I_0 \exp\left[\frac{e}{kT_c}(V_t - \Delta\phi)\right] \quad (V_t \leq \Delta\phi)$$
$$I_a = I_0 \quad (V_t \geq \Delta\phi), \quad (2)$$

where $\Delta\phi$ is the difference in work function between the target and the cathode. According to this equation the curve of $\ln I_a$ against V_t for $V_t \leq \Delta\phi$ should be a straight line with a slope of e/kT_c. However, acceptance curves of conventional camera tubes using a triode as electron gun differ from this very considerably [6]; see for example *fig. 2*. At a low potential the current accepted is much larger than the current corresponding to the Maxwellian distribution. Evidently there is a relatively large number of electrons whose velocity is high enough to enable them to reach the target in spite of the low potential. The beam temperature that can be derived from the slope of the curve for $I_a = 10$ nA is equal to about 2600 K, i.e. much higher than the cathode temperature (about 1000 K).

Why the velocity distribution of the electrons in conventional camera tubes using a triode electron gun differs so greatly from the distribution corresponding to the cathode temperature had long been a subject of investigation. It finally turned out that it is due to the

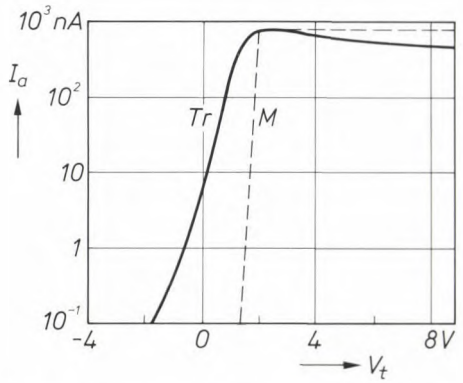


Fig. 2. Measured acceptance curve *Tr* for a triode and the curve *M*, calculated on the basis of the Maxwellian distribution of the axial electron velocity at a cathode temperature of 1000 K. The current I_a accepted by the target (SnO_2) is plotted logarithmically against the potential V_t of the target with respect to the cathode. The current with the triode is much greater than desirable, especially at a low potential.

electron interactions that can occur in a triode [7]. The cathode, the grid and the anode of a triode form a lens that focuses the electron beam delivered by the cathode into what is termed a 'crossover'. This is illustrated schematically in *fig. 3* for a conventional triode with a grid potential of −40 V and an anode potential of 300 V. In the proximity of the crossover there is a high concentration of electrons. The maximum current density at the axis is about 15 times greater than at the centre of the cathode. In addition these electrons have a relatively high energy there because the potential in this region is already fairly high (about 150 V). This concentration of high-energy electrons is accompanied by strong interaction, as a result of which some of the electrons receive an excessively high axial velocity.

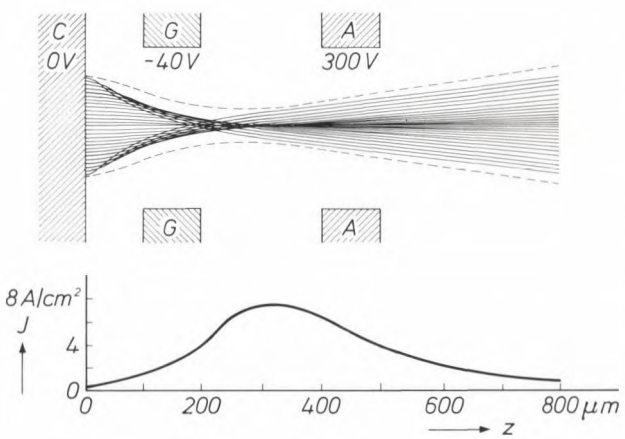


Fig. 3. *Above:* Schematic diagram of electron trajectories in a triode in which the potentials with respect to the cathode *C* are −40 V for the grid *G* and +300 V for the anode *A*. The trajectories shown are in fact the central axes of the individual beams. The dashed lines indicate the diameter of the total beam that corresponds to a current density of about a quarter of that at the axis of the tube. Strong focusing occurs in the triode, resulting in a 'crossover'. *Below:* Current density *J* at the tube axis, as a function of the position *z* on the axis. The current density at the crossover is about 15 times higher than at the cathode.

We found that a good answer to this problem was to replace the triode as electron gun by a diode with a low anode potential. *Fig. 4* shows schematically a number of electron trajectories for a diode whose anode potential is only 30 V. The effect of the electron interaction on the beam temperature is much weaker than in the triode of fig. 3, because no focusing takes place, so that there is no crossover, and because the electrons have a much lower energy owing to the much lower anode potential.

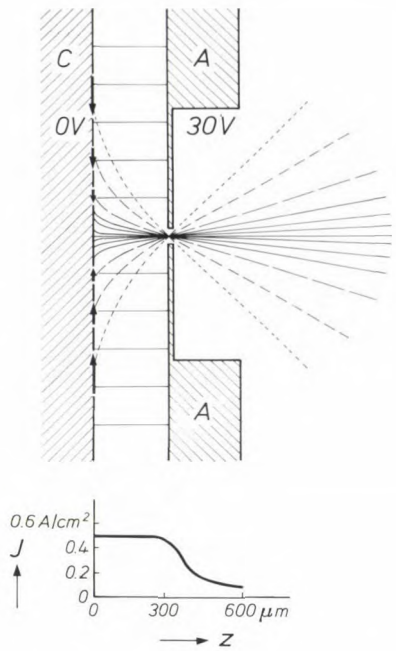


Fig. 4. *Above:* Schematic diagram of electron trajectories in a diode in which the anode *A* has a potential of 30 V with respect to the cathode *C*. An important difference from the triode of fig. 3 is that electrons with no transverse velocity at the cathode move almost parallel to the tube axis. Since the aperture in the anode diaphragm is very small, only electrons with a very selective transverse velocity can pass through it. The dashed lines indicate the trajectories of a few electrons that have no axial velocity at the cathode (*vertical arrows*), and contribute to the divergence of the emergent beam. *Below:* Current density *J* at the tube axis, as a function of the position *z* on the axis. The current density in the diode, unlike that in the triode, is nowhere greater than at the cathode.

[4] The possibility of using a diode as an electron gun in camera tubes was described by the author in Netherlands patent application No. 7013098 (1970), issued in the United States under Patent No. 3.831.058 (1974).

[5] A more detailed description is given in J. Hasker, IEEE Trans. **ED-18**, 1075, 1971.

[6] L. J. van de Polder, Philips Res. Repts **22**, 178, 1967.

[7] The effect of electron interactions on the velocity distribution has been described by K. H. Loeffler, Z. angew. Phys. **27**, 145, 1969. On the basis of this description, velocity distributions were calculated for triodes, and these agree well with the measured distributions. Basic calculations were made by E. T. Ferguson, formerly with Philips Research Laboratories.

The better velocity distribution of the electrons in the diode is clearly illustrated by the measured acceptance curve; see *fig. 5*. With decreasing target potential the accepted current in the case of the diode decreases much more strongly than with the triode. The slope of the curve indicates a good approximation to the Maxwellian distribution corresponding to the cathode temperature. The electron temperature derived for $I_a = 10$ nA is 1185 K, which, compared with the value found for the triode (about 2600 K) is a much smaller deviation from the cathode temperature (about 1000 K).

The divergence of the electron beam emerging from the anode diaphragm of the diode is much greater at 30 V than at 300 V, for example; see *fig. 6*. The cause of the divergence is the thermal spread in the transverse velocities of the electrons that leave the cathode surface (fig. 4). In the very small aperture in the anode diaphragm the current density is homogeneous. If the lens action of this aperture is neglected, the 'brightness distribution' $f(\alpha)$ for the divergence of the transmitted electrons can be assumed to be Gaussian. Expressing $f(\alpha)$ as a current per unit area and per unit solid angle, it is given by

$$f(\alpha) = \frac{J_0}{\pi\alpha_{1/e}^2} \exp(-\alpha^2/\alpha_{1/e}^2), \qquad (3)$$

where α is the angle to the tube axis. J_0 is the axial current density at the location of the diaphragm and can be set equal to the current density at the cathode (fig. 4). The angle $\alpha_{1/e}$ corresponds to the (1/e)-value of this brightness distribution. The value of $\alpha_{1/e}$, a measure of the divergence, depends on the cathode temperature T_c and the anode potential V_a:

$$\alpha_{1/e} = \sqrt{kT_c/eV_a} \text{ radians.} \qquad (4)$$

At a normal cathode temperature of 1000 K, $\alpha_{1/e}$ is equal to 3° if $V_a = 30$ V and less than 1° if $V_a = 300$ V.

The divergence of the electron beam from a diode with a low anode potential presents a problem if it is desired to focus the beam with minimum aberration. For electrostatic focusing in camera tubes a unipotential lens is generally used, consisting of three electrodes, with the middle one at a relatively low potential (e.g. 50 V) and the other two at the same high potential (e.g. 300 V). In our case, however, a unipotential lens is not suitable, because the anode potential of the diode is too low and the beam too divergent. To obtain the desired electrostatic focusing of the beam in spite of this, we decided to use an accelerating lens consisting of an electrode at a low potential connected to the anode of the diode, and an electrode at a high potential.

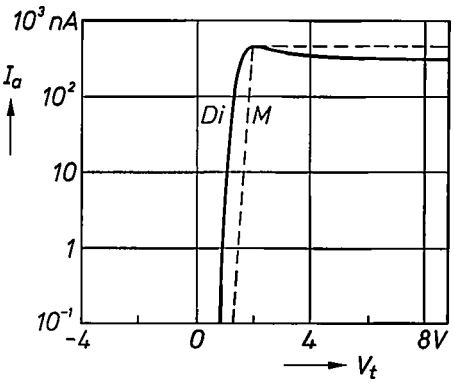


Fig. 5. Measured acceptance curve *Di* for the diode in fig. 4, and the calculated Maxwellian curve *M*. The measured curve indicates a velocity distribution that approximates closely to the Maxwellian distribution, and is much better than that in the triode.

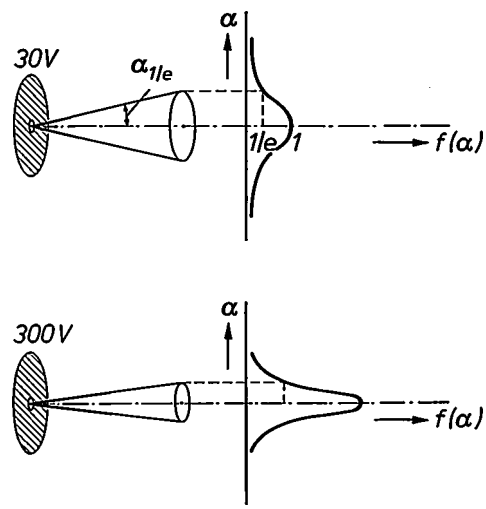


Fig. 6. Divergence of the electron beam emerging from an aperture in the anode diaphragm of a diode for anode potentials of 30 V and 300 V. The angle $\alpha_{1/e}$ corresponds to the (1/e)-value of the distribution function $f(\alpha)$, where α is the angle to the tube axis. At 30 V, $\alpha_{1/e}$ is about 3°, at 300 V it is less than 1°.

Focusing with an accelerating lens

Fig. 7 shows a diagram of focusing by a unipotential lens and also by an accelerating lens. Both lenses are provided with a diaphragm to reduce the spherical aberration. The electron trajectories, and hence the focusing, are determined by the location of the equipotential surfaces. In the unipotential lens the divergent beam enters a region of a decreasing potential, so that the divergence at first becomes even greater there. Next, as the beam enters the region of increasing potential, it starts to converge, strongly at first and then more weakly, until there is focusing at the axis. The accelerating lens has a potential variation along its axis that corresponds to that of the second part of the unipotential lens, so that the operation is convergent all the way from the start to the focusing. In

the accelerating lens the diameter of the comparable part of the beam transmitted by the diaphragm is therefore smaller at the centre of the lens than in the unipotential lens. The result of this is that there is less aberration in the accelerating lens, since the beam is narrower during the focusing, and better resolution can therefore be obtained.

The ratio of the current I_2 passing through the diaphragm D_2 of the accelerating lens to the current I_1 passing through the anode diaphragm D_1 is given by

$$I_2/I_1 = 1 - \exp(-r_2^2/r_{1/e}^2), \qquad (5)$$

where r_2 is the radius of the aperture in D_2, and $r_{1/e}$ is the height at which an electron starting from D_1 with an angle $\alpha_{1/e}$ (eq. (4)) arrives at the second diaphragm D_2. For r_2 we choose a value such that the spherical aberration has an acceptable effect on the sharpness with which the aperture in D_1 is imaged on the photoconducting layer [8]. The loss of beam current can then be limited by ensuring that $r_{1/e}$ is small. For this reason the lens with D_2 is located not too far away from D_1. The high potential of D_2 then ensures that the beam is focused in such a way that the current density is high enough at D_2.

The magnification with which the aperture in D_1 is imaged on the photoconducting layer is similar to the magnification that would be obtained by ordinary optical means; see *fig. 8*. Instead of the light velocities in the two media, expressed in refractive indices, we now have the electron velocities, which are proportional to the square root of the initial potential V_i and the square root of the final potential V_f. For the paraxial magnification M we can use the following expression:

$$M = \frac{b}{a}\sqrt{V_i/V_f}, \qquad (6)$$

where a and b are the object and image distances, as shown in fig. 8. To obtain sharp images a small M is required. With a camera tube of given length the ratio b/a can be made smaller by increasing the distance between D_1 and the lens with D_2. As we have seen, however, this entails a greater loss of beam current. An acceptable compromise implies making b/a roughly equal to 2.5. For comparison, when a unipotential lens is used, a and b are roughly the same. The final potential in an accelerating lens can be so much greater than the starting potential, however, that the magnification it gives is nevertheless substantially less than that given by a unipotential lens. Very sharp images can therefore be obtained by using an accelerating lens with a diaphragm.

Another advantage of this lens is that the aperture in the anode diaphragm D_1 of the diode does not need to be aligned with extreme accuracy in relation to the tube axis. This is because the small aperture in the anode diaphragm functions as a point source with a relatively large beam angle. This implies that, to a first approximation, it makes little difference exactly which part of the beam passes through the second diaphragm D_2. A transverse displacement of the point source has hardly any effect on the operation of the tube, so that it would even be permissible to allow some eccentricity.

Another important advantage is that the emission of electrons is hardly affected at all by positive ions, because only few ions can bombard the cathode at the location where the electrons originate. In the focusing region of the accelerating lens the potential is so high that ions are easily formed. They have very little chance of bombarding the cathode, however, because they hardly ever come through the aperture in D_1,

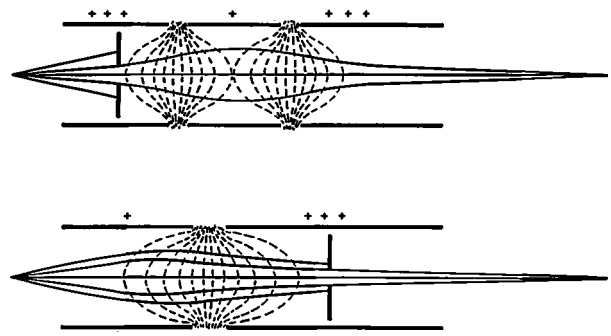


Fig. 7. Equipotential surfaces (*dashed lines*) and electron trajectories (*solid lines*) in a unipotential lens (*above*) and in an accelerating lens (*below*). + low potential. +++ high potential. Both lenses have a diaphragm. The electrons are subject to a force acting perpendicular to the local equipotential surface, in the direction of increasing potential. In the unipotential lens the beam first becomes more divergent, then becomes convergent again and is focused at the axis. In the accelerating lens (corresponding to the second part of the unipotential lens) the electrons encounter a steadily increasing potential, so that they are continuously focused towards the axis right from the start.

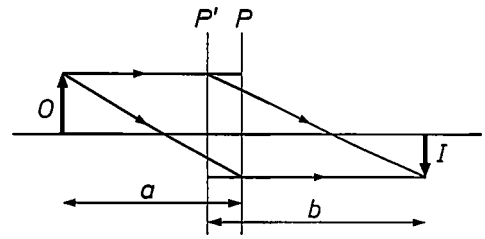


Fig. 8. Imaging with an accelerating lens. As in light optics, the image I of a particular object O can be derived from the object principal plane P at a distance a from the object, and the image principal plane P' at a distance b from the image. The ratio of the dimensions of I to those of O is the paraxial magnification M, given in eq. (6).

[8] J. H. T. van Roosmalen, IEEE Trans. ED-18, 1087, 1971.

starting as they do with a velocity of nearly zero and then moving in directions virtually perpendicular to the equipotential surfaces of the accelerating lens and eventually landing on the wall and on D_1; see *fig. 9*.

The use of an accelerating lens with diaphragm also has advantages in connection with the magnetic deflection of the electrons. The electrons pass through the second diaphragm to enter a region where the potential varies only very little with location ('equipotential region'). When the electrons are magnetically deflected in this region, there is hardly any interaction between the focusing and the deflection, so that these can be controlled and corrected independently of each other. This is one of the reasons why it is possible to obtain exceptionally good registration (superposition) of the three images with a colour-television camera using three camera tubes of the new design.

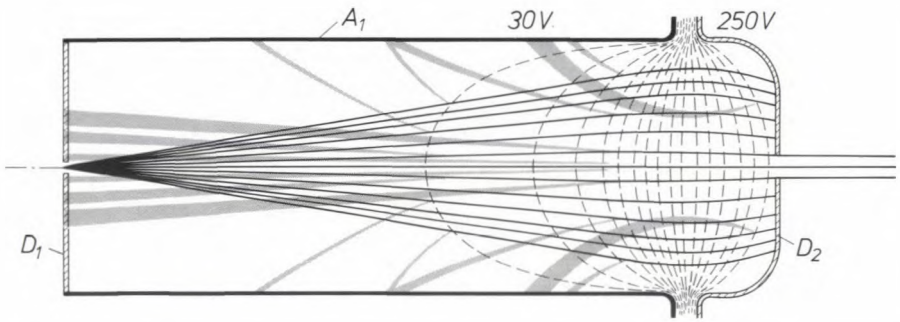


Fig. 9. Schematic cross-section of an accelerating lens, consisting of the anode A_1 with diaphragm D_1, at a low potential (30 V), and a second diaphragm D_2, at a high potential (250 V). The solid lines are the electron trajectories and the grey beams represent the trajectories of the ions. Positive ions formed in the lens have initially almost zero velocity and therefore move almost perpendicular to the equipotential surfaces (*dashed lines*) in the lens. For that reason only few ions pass through the aperture in D_1.

After magnetic deflection the electrons no longer move parallel to the tube axis, of course. This must, however, be the case when they land on the photoconducting layer, because oblique landings can give rise to undesirable effects [2]. For this reason we use a landing-correction lens, rather like a second accelerating lens with the diaphragm D_2 as the initial electrode. The final electrode of this lens consists of a ring with a mesh at a high potential; see *fig. 10*. The lens action has the effect of deflecting the electrons back to such an extent that at the mesh they are again moving parallel to the tube axis and are then able to make an almost perpendicular landing. The strongest possible lens is required: with a low anode potential (typically 250 V) in the deflection region the deflection energy needed is low, whereas with a high mesh potential (typically 750 V) the accelerating lens has a high final potential, so that, as eq. (6) shows, the total magnification is small. Another advantage of a high mesh field between the mesh and the photoconducting layer, give rise to an undesired change of potential ΔV ('landing error') on the layer, which is greater the higher the mesh potential V_m [2]:

$$\Delta V = V_m \sin^2 \beta. \qquad (7)$$

Furthermore, the barrel distortion accompanying the trajectory correction of the landing-correction lens increases with the potential ratio of its electrodes. This implies that the lens must be made very accurately; with the new technology that will now be described this is readily possible.

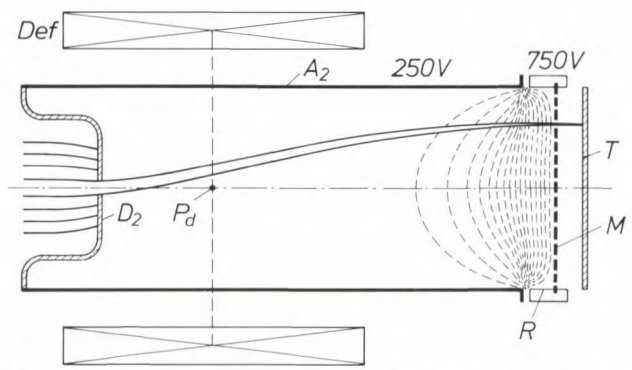


Fig. 10. Schematic cross-section of a landing-correction lens, consisting of an anode A_2 with the diaphragm D_2, at 250 V, and a ring R with a mesh M at a much higher potential (750 V). The lens causes the electrons, deflected by the deflection coils *Def*, to land perpendicularly on the photoconducting layer T. P_d deflection point, where the deflected electrons apparently originate.

The new technology

It was necessary to look for a new technology because the new electron-optical possibilities could not be fully realized with the conventional technology. In the conventional technology the tubular electrodes are mounted on a jig and attached to each other by means of insulating rods. This assembly is then fused into a glass envelope which is subsequently evacuated. The deflection coils are fixed to the outside wall by means of a coil sleeve. Camera tubes produced in this way occasionally present problems of registration in a colour-television camera. One of the reasons for this is that the axis of the electrode system is only indirectly correlated with that of the coils. The insulating rods take up a relatively large amount of space, and the residual gas pressure is adversely affected by the large total surface of the electrodes.

In the new technology, steps in the inside diameter of a glass envelope serve as locating faces on which to mount electrodes in the tube. The other electrodes are applied as a thin metal film to the inside wall; this gives considerable space saving. The deflection coils also take up less space: because of the highly accurate dimensions of the envelope, the coils can be fitted directly to the outside wall, without an internal coil sleeve. The smaller coil diameter also means a lower deflection energy. *Fig. 11* represents schematically two camera tubes both of which are based on the new electron-optical design, but with one made in the conventional way and the other by the new technology. We shall now take a closer look at the new technology.

The glass envelope with the specific shape required is made by means of the vacuum-forming technique. In this technique a pre-formed glass tube is placed on a metal mandrel. The tube with mandrel is evacuated and heated to the softening temperature of the glass. Because of the external air pressure the softened glass is forcibly pressed against the mandrel. In this way the glass envelope is given a highly accurate and reproducible imprint of the profile of the mandrel; see *fig. 12*. Upon cooling to room temperature the metal shrinks more than the glass, so that the mandrel works loose and can now be separated from the profiled envelope. Measurements on a large number of tubes have shown that the internal dimensions of the glass envelopes successively formed in this way upon one mandrel are identical with each other to within less than 2 μm.

To permit electrical connections to the electrodes that still have to be fitted, the envelope has to be provided at various points with lead-throughs. For this purpose holes are made in the glass wall. To produce a thin metallic film on the inside wall the glass envelope is first metallized over its entire surface,

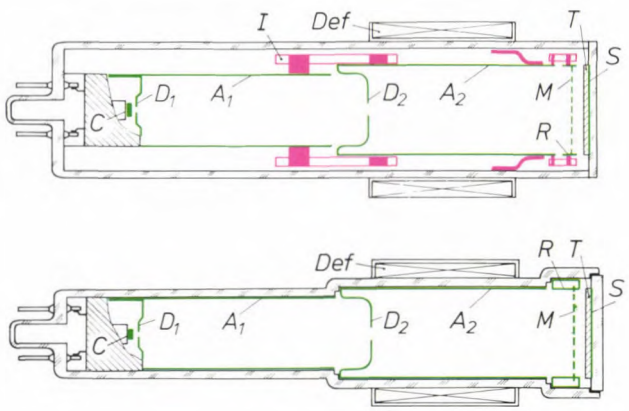


Fig. 11. Schematic cross-section of a camera tube based on the new electron-optical design, made with the conventional technology (*above*) and with the new technology (*below*). *C* cathode. A_1 first anode with diaphragm D_1. A_2 second anode with diaphragm D_2. *Def* deflection coils. *M* mesh. *R* ring. *T* target (photoconducting layer). *S* signal plate. In the conventional technology the electrodes (*green*) are mechanically mounted by insulating rods *I* (*red*), and the small springs (*red*) are used to align the electrodes inside the envelope; an internal coil sleeve is also required. In the new technology all this is unnecessary: some parts of the electrodes are applied as a metal film to the inside of the envelope and the coils can be located directly on the outside of the envelope.

Fig. 12. *Above:* Photograph of a glass envelope, vacuum-formed on a profiled metal mandrel, and the mandrel, inside a half-section of the envelope. *Below:* A detail to illustrate the very close fit.

including the lead-through holes. This is done by means of 'electroless' metallization, a wet-chemical process in which the whole glass surface is coated with a metal film. At the places where no metal is required, on the whole of outside wall and a small part of the inside wall, the film is etched away. The film in and around the lead-through holes is maintained by applying a protective layer of resist before etching. The

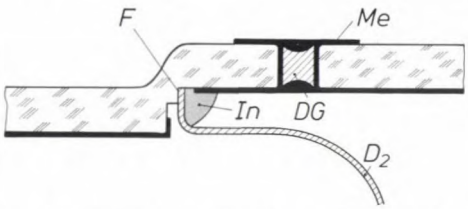


Fig. 13. The electrical connection and the method of attaching the second diaphragm D_2 in the new technology. A metal film *Me* is applied not only to the inside wall of the envelope but also in and around the lead-through hole, and the hole is then made vacuum-tight by sealing with devitrifying glass *DG*. The diaphragm is fitted against the locating face *F* in the glass envelope. It is then soldered to the metal film by means of small indium beads *In* and thus electrically connected to it at a position where this does not perturb the electron optics.

Fig. 14. Photograph of an experimental Plumbicon tube in a coil unit, made with the new technology on the basis of the new electron-optical design.

required insulation between the electrodes A_1 and A_2 and between A_2 and the mesh electrode M (fig. 11) is obtained by grinding away the metal film at these locations.

The electrical connection to the electrodes is made by means of the metallized wall of the lead-through holes; see *fig. 13*. The holes are sealed vacuum-tight by means of devitrifying glass. This is heated to the molten state and then allowed to cool, in which process it crystallizes to form the vacuum-tight seal. The temperature required for this seal lies below the softening temperature of the glass, so that the accuracy of the glass envelope is not affected.

Next the diaphragm D_2 and the mesh M are mounted on the ledges on the inside wall of the glass envelope. These components are fixed in position by applying small beads of indium that melt during the outgassing and bake-out. The electrodes are thus soldered to the metal film and electrically connected to it (fig. 13). Indium has a number of suitable properties for connection of the electrodes: it has a low vapour pressure, good adhesion, a low melting point and is very ductile, so that no stresses occur in the glass envelope when the tube is cooled down from the temperature required for outgassing.

The diode gun is mounted and then slid into the envelope and secured at the desired position. Then the window with the signal plate and the photoconducting layer are fitted, with the aid of a thin ring of indium laid around the edge of the tube head. The window is then pressed into the tube head by means of the indium seal. Because the seal is so very thin, the distance between mesh and layer is accurately reproducible. To minimize electrical noise, the capacitance of the signal plate to the surrounding earthed parts must be kept as small as possible. The electrical connection to the signal plate is therefore made via a lead-through in the window. The lead-through hole is sealed vacuum-tight in the same way as the other lead-through holes. The tube base is fixed to the tube wall by means of devitrifying glass, and the tube is then evacuated.

Implementation in camera tubes

The new electron-optical design has been put into hardware by means of the new technology in a number of experimental Plumbicon tubes. *Fig. 14* shows a photograph of such a camera tube with its coil unit. A tube of this type can be made extremely accurately. It is much more compact than other types of camera tube; see *fig. 15*. The characteristics of the experimental tubes are promising.

Because of the low electron temperature (fig. 5) the speed of response is relatively high. Measurements of the speed of response have been made by H. van Tongeren of the Philips Electronic Components and Materials Division. *Fig. 16* shows a plot of the signal current, measured as a function of time, for an experimental tube at a sudden transition from dark to light and vice versa. The rise lag and decay lag are very small: after 60 ms (3 fields) the signal current has increased to about 95% of the maximum value or fallen to about 5%, as appropriate.

Owing to the use of accelerating lenses with a low initial potential and a high final potential, the paraxial magnification of the total system is only 0.5, which is much smaller than is usually possible with other types

(0.7 to 0.8). This makes it possible to obtain very sharp pictures, even when the surface of the scanned photoconducting layer is very small (e.g. with a diagonal of 8 mm). The sharpness of the pictures is found from measurements of the modulation depth. In these measurements the electron beam makes a horizontal scan of an alternating charge pattern obtained from black and white stripes of different widths [1]. The fundamental frequency of the signal thus obtained is 0.5 MHz for the widest stripes. The amplitude of the signal decreases as the number of stripes per unit length (the spatial frequency) becomes larger. The modulation depth m is defined as the ratio of the amplitude at a particular frequency to the amplitude at 0.5 MHz. *Fig. 17* shows a curve of the measured modulation depth for an experimental tube

Fig. 15. Photograph of the experimental camera tube in fig. 14 and three other types.

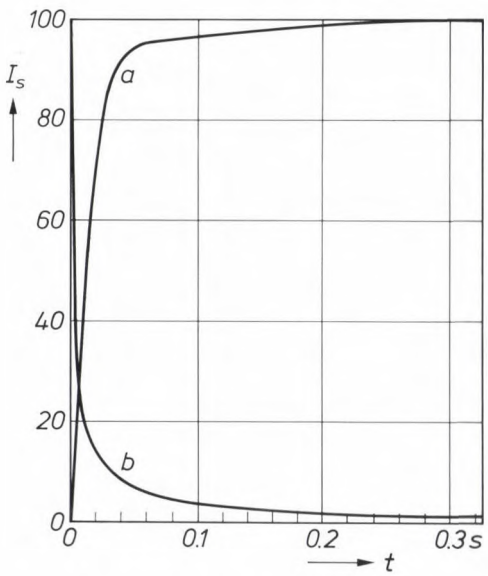


Fig. 16. The measured signal current I_s (relative units) for an experimental camera tube as a function of time t, for a sudden transition from dark to light (*a*) and vice versa (*b*). A period of 20 ms corresponds to one field. Both rise lag and decay lag are very small. The measurements were made at a beam current of 200 nA and a maximum signal current of 20 nA. The dark current of the tube was 0.5 nA and the layer capacitance was 300 pF.

plotted against the spatial frequency f. The modulation depth for the new tube is found to be very high. At the corners, even without any additional focusing the modulation depth is not much lower than at the centre of the screen.

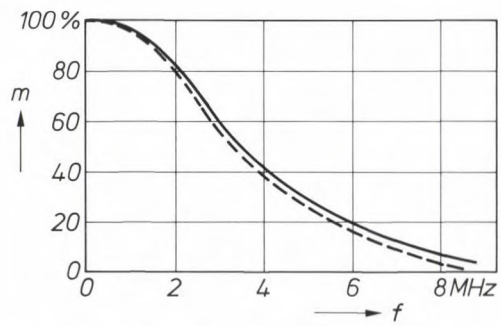


Fig. 17. Measured modulation depth m, as a function of the spatial frequency f, for an experimental tube. The modulation depth, a measure of the image sharpness, is very high, not only at the centre of the target (*solid line*) but also at the corners, without additional focusing (*dashed line*). The measurements were made with white light at a beam current of 200 nA and a signal current of 100 nA.

When three experimental camera tubes were used in a colour-television camera, accurate and stable registration was obtained. The landing errors and distortion of the raster that occur during scanning of the photoconducting layer can be made negligibly small. *Fig. 18* gives for various values of the mesh potential the measured landing error ΔV as a function of the distance r to the centre. At the scanning diagonal of 8 mm (r maximum 4 mm) the landing error on the whole surface to be scanned is very small, at a mesh potential of 700 to 800 V.

The deformation of the rectangular image, the 'distortion', can be kept small because the deflection coils also give a certain amount of distortion, which has a sign opposite to that of the landing-correction lens. *Fig. 19* shows the measured distortion D for an experimental tube with coils as a function of the mesh potential V_m. With increasing mesh potential the distortion is first negative and goes positive at about 600 V. At 700 to 800 V, the optimum mesh potential for a very small landing error, the distortion is small, only 0.3 to 0.9%. After optimization of the design and position of the deflection coils, the distortion will be even smaller.

A further important feature is that residual gases in the tube have little effect on the performance, because positive ions are practically unable to reach the cathode (fig. 9). Confirmation of the insensitivity to gases has come from experiments by P. P. M. Schampers of these laboratories. He filled a large number of tubes with different gases at a pressure of 10^{-2} to 10^{-1} Pa. During a period of more than 2000 hours there were no irregularities at all in the electron emission of the cathode. Finally, we should mention the advantage that no external provisions are required for aligning the electron beam with the tube axis.

The investigation described here was carried out in close cooperation with G. A. H. M. Vrijssen. Major contributions were also made by A. H. M. de Gouw, P. F. A. Haans, F. C. M. de Haas, J. P. Hornman, M. A. Karsmakers, P. Lavrijsen and H. F. Wilbrink. The coil design was worked out in cooperation with P. H. Carboex and H. J. van Rooy of the Philips Video Division.

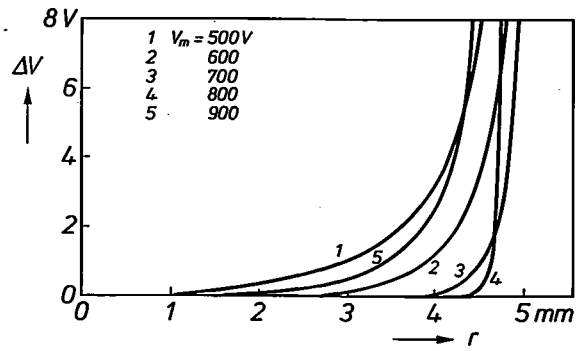


Fig. 18. Measured landing error ΔV in an experimental tube as a function of the distance r to the centre of the photoconducting layer, for different values of the mesh potential V_m. The best value for the mesh potential in this case is about 700 to 800 V. On a scanning diagonal of 8 mm, i.e. $r \leq 4$ mm, the landing error is very small.

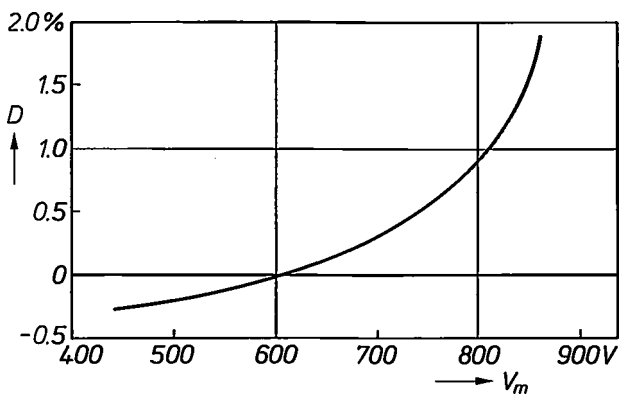


Fig. 19. The distortion D, relating to the height of a rectangular image, measured in an experimental tube as a function of the mesh potential V_m. Even at a high mesh potential of 700 to 800 V the final distortion is small.

Summary. In the new electron-optical concept for television camera tubes the electron beam for scanning the photoconducting layer is provided not by a triode but by a diode with a relatively low anode potential. The divergent beam coming from the anode diaphragm is electrostatically focused by means of an accelerating lens. The electrons are magnetically deflected, and a landing-correction lens ensures that they land perpendicularly on the photoconducting layer. The design has led to the development of a new technology for making camera tubes. In this technology some of the electrodes are applied in the form of a thin metal film on the inside wall of a glass vacuum envelope, while the others are accurately positioned by means of locating faces made in the envelope. The design and the technology make it possible to produce compact camera tubes that have good characteristics with regard to speed of response, image sharpness and registration.

A new generation of high-pressure sodium lamps

C. A. J. Jacobs and J. A. J. M. van Vliet

Although existing high-pressure sodium lamps (SON) have a very high luminous efficacy, Philips continue to look for ways of improving their performance. Research on the high-pressure sodium discharge — which has already been reported in detail in the special 'Light' issue of this journal in 1975 — has enabled us to design a new generation of lamps, which differ very little from existing lamps in most of their characteristics — so that they can replace lamps in existing installations — but have a luminous efficacy about 15% higher. One type (150 W) is already in production, a second (50 W) is to follow shortly and several other types are under development.

Introduction

Lamps in which the light is generated by means of a high-pressure sodium discharge contain sodium vapour, a little xenon to ensure that the cold lamp ignites and the sodium discharge comes into operation, and mercury vapour at a relatively high pressure. By adding this heavy gas — the buffer gas — it is possible to give the discharge the electrical and spectral characteristics that are desirable in a lamp that is to combine a high luminous efficacy with a convenient size and an acceptable chromaticity and can be run economically from the lighting mains [1].

It has been known for some years that it would be more advantageous for the efficacy if xenon could be used for the buffer gas instead of mercury vapour [2], but for various reasons manufacturers were not able to do this. Replacing the mercury by xenon would result in an unsatisfactory colour, a longer discharge tube and, finally, an extremely high ignition voltage. It was this high voltage that appeared to be the insuperable problem. Research has shown that the first two objections can be overcome, and the advantage of the better efficacy retained, by adding a certain amount of mercury to the xenon. It has also been found that the ignition voltage can be reduced below the limit specified in the draft standards now in circulation, by fitting a simple auxiliary electrode around the discharge tube [3].

To give some idea of the nature of these problems and how serious they are, we shall first consider how the efficacy, chromaticity and the electrical characteristics of a high-pressure sodium discharge depend on the nature and pressure of the buffer gas. The final section will summarize the characteristics of three of the new generation of lamps: the 150 W lamp already in production, and experimental versions of 70 W and 400 W lamps.

Luminous efficacy and chromaticity

The luminous efficacy η of a light source is proportional to the fraction of the electrical power supplied that is converted into visible radiation, multiplied by the luminous efficiency V_s of the visible radiation produced:

$$\eta = 683 \frac{P_{vis}}{P_{in}} V_s \text{ lm/watt.}$$

C. A. J. Jacobs and Ir J. A. J. M. van Vliet are with the Philips Lighting Division, Eindhoven.

[1] See for example J. J. de Groot, J. A. J. M. van Vliet and J. H. Waszink, Philips tech. Rev. 35, 334, 1975.
[2] K. Schmidt, C. R. 6e Conf. int. sur les Phénomènes d'ionisation dans les gaz, Paris 1963, Vol. III, page 323.
[3] I. Iwai, M. Ochi and M. Masui, J. Light & vis. Env. 1, No. 1, page 7, 1977.
C. A. J. Jacobs, L. Sprengers and R. L. C. de Vaan, J. Illum. Engng Soc. 7, 125, 1978.

The ratio P_{vis}/P_{in} (the radiant efficiency) is not very different for mercury and xenon. *Fig. 1* indicates how the radiant efficiency for the two buffer gases depends on their pressure p_b. This pressure has been expressed as a multiple of the pressure p_{Na} of the sodium vapour. In all our experiments the pressure p_{Na} had the value p_{10}, at which the separation of the two maxima in the self-reversed D line is 10 nm; at this pressure η has the highest value for a discharge in pure sodium vapour [1]. The advantageous effect of the buffer gas is unmistakable, but it is also clear that there is no point in making p_b/p_{Na} larger than 20 to 30.

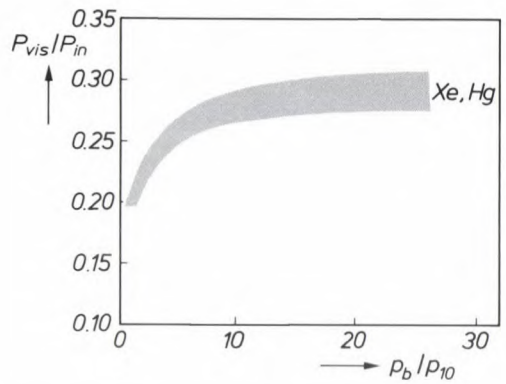


Fig. 1. The radiant efficiency, i.e. the ratio of the power P_{vis} radiated in the visible region to the electrical power supplied P_{in}, as a function of the pressure p_b of the buffer gas. The pressure is expressed as a multiple of the value p_{10} of the sodium vapour pressure when the maxima of the self-reversed sodium D line differ by 10 nm (about 15 kPa) [1][4]. The discharge took place in a tube with an inside diameter of 4.8 mm and a wall temperature of 1500 K (at the hottest point), values that apply for the discharge tube of a 150 W SON lamp. At a buffer-gas pressure of about $20p_{10}$ or higher the radiant efficiency is much higher than at low pressure, but it makes little difference whether mercury vapour or xenon is used — the difference is less than the accuracy of the measurement.

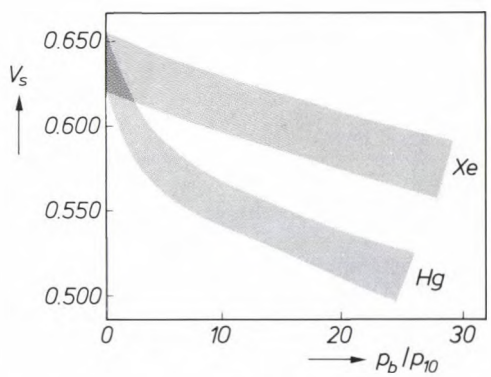


Fig. 2. The luminous efficiency V_s of a high-pressure sodium discharge — the conditions are the same as in fig. 1 — again as a function of p_b/p_{10}. Increasing the buffer-gas pressure gives a reduction in V_s, but this effect is much weaker for xenon than for mercury.

The ratio p_b/p_{Na} is a measure of the ratio of the buffer gas to the sodium vapour, and hence of the thermal conductivity of the mixture. As p_b/p_{Na} increases, the thermal conductivity decreases; this is an important quantity for the designer [1].

The advantages of using xenon as the buffer gas instead of mercury come about because of the different effects the two gases have on the spectral energy distribution and hence on the luminous efficiency V_s. Both gases reduce the luminous efficiency V_s as compared with pure sodium vapour, but this is much smaller for xenon than for mercury; see *fig. 2*. An examination of the spectra (*fig. 3*) reveals that as a result of the addition of each of the gases the red 'wing' (the longer-wave side) of the broadened sodium doublet is raised; this results in a reduction of V_s because the eye is not very sensitive in this wavelength range and the power is taken from the range where the eye is more sensitive. This effect is most marked in mercury, and in addition the spectrum for mercury has two bands in this region (at 655 and 670 nm). In xenon, on the other hand, there is a strong band at about 560 nm just where the eye is very sensitive.

Obviously the changes produced in the spectrum of the discharge as a result of the addition of mercury or xenon as the buffer gas have some effect on the colour of the emitted light. As *fig. 4* shows, the addition of mercury improves the spectrum at first — the light source becomes more like a black-body radiator — but when the mercury pressure is increased further the colour coordinates of the source move towards the red. The addition of xenon makes it move towards the yellow-green and this is less desirable.

Electrical characteristics

One of the advantages of using mercury as the buffer gas is that the field-strength in the discharge is higher, which means that for a given operating voltage the discharge tube can be shorter — and thicker — than for a discharge in pure sodium vapour. Xenon does not offer this advantage; the addition of xenon to the sodium vapour has very little effect on the field-strength. To obtain a reasonable operating voltage it would therefore be necessary to make the tube longer but, to keep the 'wall loading' [1] at the same value there would have to be a simultaneous reduction in the diameter. However, this has an adverse effect on P_{vis}/P_{in}. A lamp like this would also

[4] J. J. de Groot and J. A. J. M. van Vliet, Determination of the sodium and mercury vapour pressure in high pressure sodium lamps, in: 2nd Int. Symp. on Incoherent light sources, Enschede 1979, page 30.
J. J. de Groot and J. A. J. M. van Vliet, J. Physics D **8**, 651, 1975.

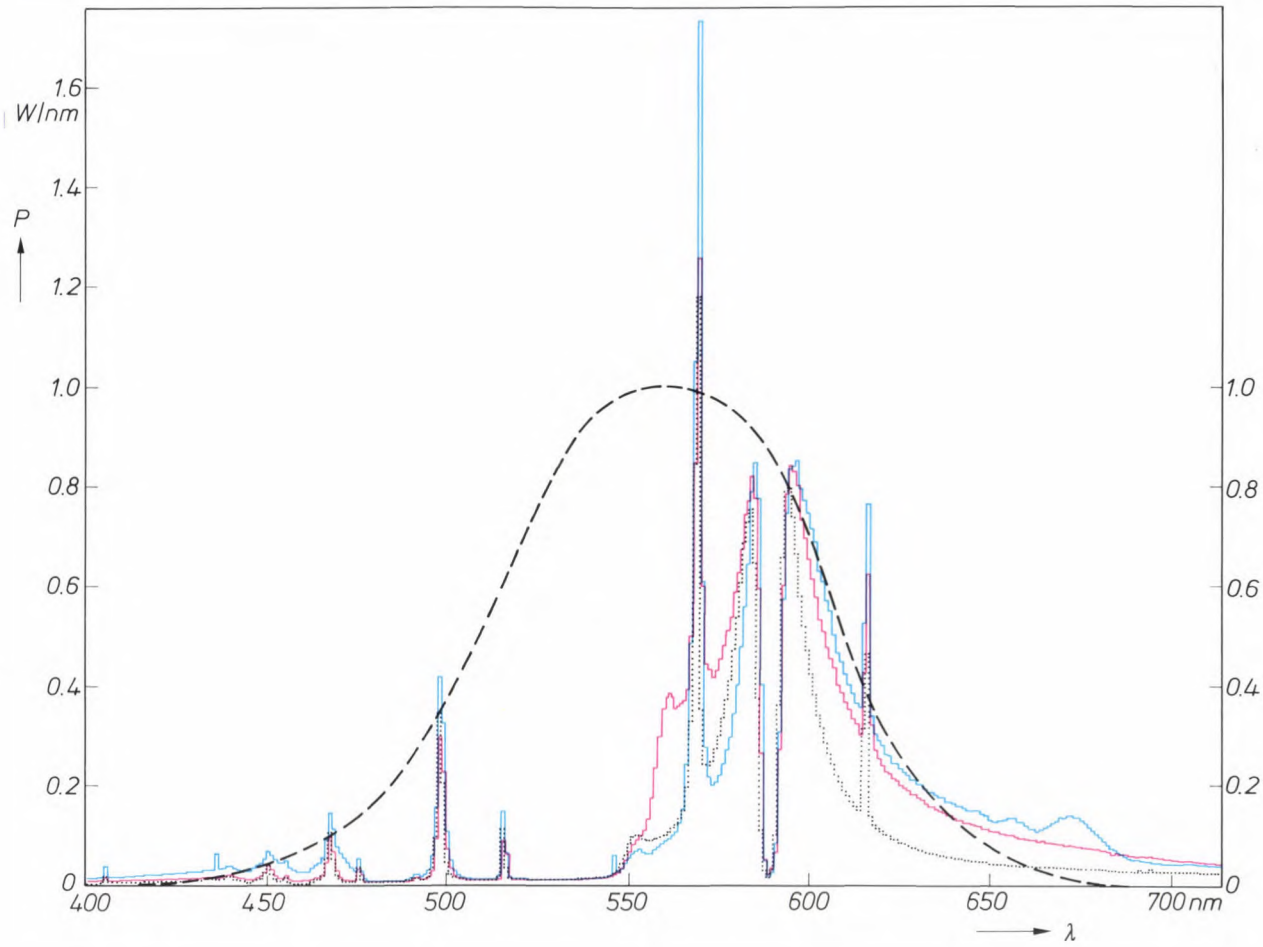


Fig. 3. The spectrum — power P per nm plotted against the wavelength λ — of high-pressure sodium discharges with mercury vapour (*blue*) and xenon (*red*) as the buffer gases. The dotted line gives the spectrum of pure sodium vapour. All the spectra relate to the sodium pressure p_{10} and are absolute spectral power distributions for an input power of 150 W. The dashed curve is the spectral luminous efficiency (see the scale on the right). The spectrum of the discharge with xenon as the buffer gas is clearly more suitable than the spectrum of a discharge with mercury as the buffer gas because the red 'wing' is lower, mainly as a result of the presence of a strong band at about 560 nm.

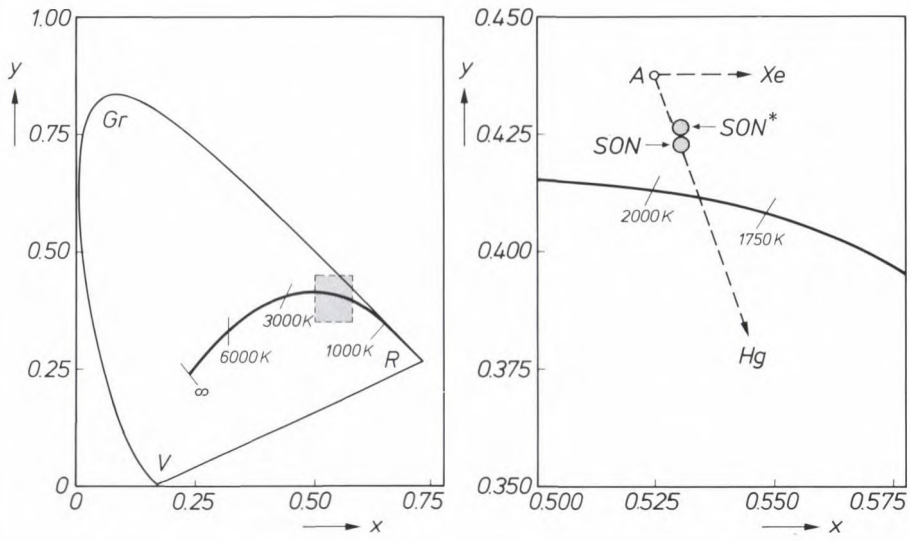


Fig. 4. Chromaticity diagram (*a*) with detail (*b*), illustrating the chromaticity of the high-pressure sodium discharge under different conditions. *Gr*, *R* and *V* indicate the green, red and violet corners of the colour triangle. *A* discharge in pure sodium vapour (no buffer gas). Adding mercury vapour shifts the point in the direction *Hg*, adding xenon moves it in the direction *Xe*. The solid curve gives the chromaticity of a black-body radiator — each point corresponds to a given temperature. The chromaticity coordinates of the lamps of the first generation (*SON*) are $x = 0.53$, $y = 0.42$, and the chromaticity of the new lamps (*SON**) is close to it ($x = 0.53$, $y = 0.425$).

be rather inconvenient for practical lighting fittings.

A greater disadvantage is that a very high voltage is necessary to ignite lamps in which xenon is the buffer gas. The starters that have been developed for the SON lamps cannot ignite an Na-Xe lamp of the same dimensions when the xenon pressure is higher than 6 kPa (i.e. about 50 kPa in the operating lamp) [5]. Since mercury and xenon lamps should be interchangeable and there should be no flashover in the starter or between the supply leads in the base of the lamp, this is an unacceptable situation.

The new generation of SON lamps

As already stated briefly in the introduction, the various difficulties of using xenon in the new generation of high-pressure sodium lamps have largely been overcome. It has been found that the dimensions of the discharge tube can be kept the same as those in the existing SON lamps — and almost all the advantages retained — by using a mixture of xenon and mercury as the buffer gas instead of xenon alone. This also had the important advantage that the chromaticity shifted appreciably in the desired direction to a position very close to the chromaticity of the existing lamps.

The ignition electrode that is used to overcome the starting problem is nothing more than a wire of tungsten or tantalum running the entire length of the discharge tube and connected to one of the main electrodes through a capacitor or a bimetallic strip. The wire makes starting easier because it considerably reduces the effective electrode spacing for the starting pulse. In the new 150 W lamps the wire is capacitively coupled to one of the main electrodes and is mounted so that it lies against the discharge tube; see *fig. 5*. In lamps with a bimetallic strip this element moves the wire away from the discharge tube as soon as it starts to get hot. The purpose of both of these measures is to prevent a high electrical field-strength from building up in the wall of the discharge tube. This would cause migration of sodium ions and hence local blackening of the tube.

The characteristics of three SON lamps of the new generation — the 70 W lamp and the 400 W lamp are still in the experimental stage — are set out in *Table I*. It can be seen that an improvement of 10 to more than 15% in the luminious efficacy has been obtained.

There are a number of other important improvements. The new lamps have a much shorter warming-up time than their predecessors (about a minute). Sputtering of the electrodes — in the first few moments after switching on — and hence blackening of the tube have been reduced considerably; see *fig. 6*. This sputtering effect is particularly important be-

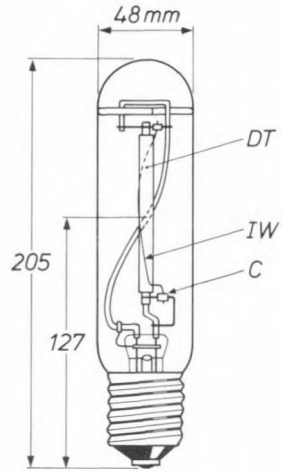


Fig. 5. Diagram of a new 150 W SON lamp with clear cylindrical envelope. (A version is also available with an oval envelope with light-diffusing coating.) *DT* discharge tube. *IW* ignition electrode, connected to one of the electrodes through the ceramic capacitor *C*.

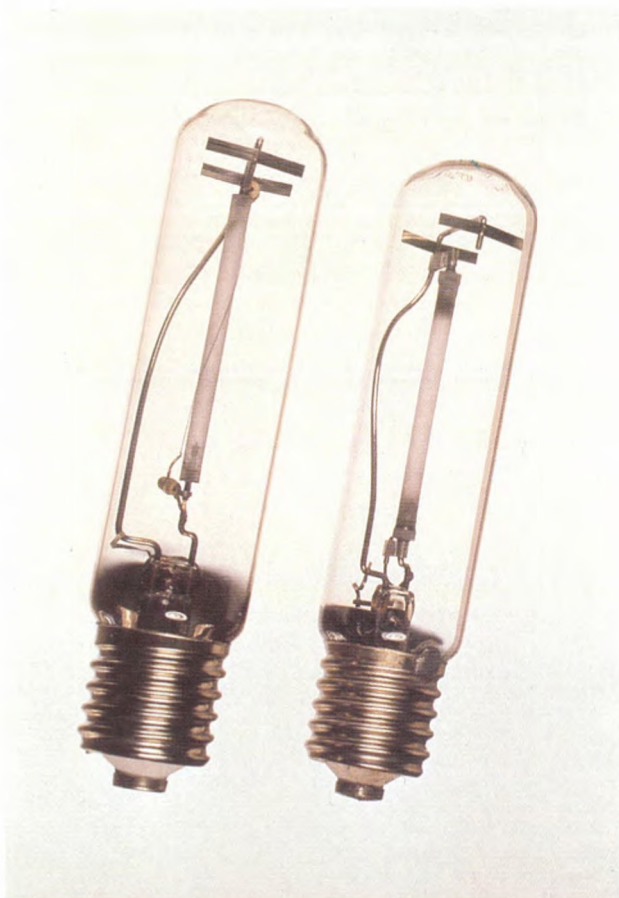


Fig. 6. A 150 W SON lamp of the first generation (*right*) and one of the new generation, both after 10 000 hours of operation. The discharge tube of the older SON lamp reveals a definite blackening near the electrodes but the new lamp does not. The effect of the difference is that the luminous efficacy of the new SON lamps has hardly decreased at all during the 10 000 hours of operation.

[5] J. A. J. M. van Vliet and C. A. J. Jacobs, New developments in high pressure sodium discharge lamps, in: CIBS Nat. Lighting Conf. 1980, Canterbury, page CA1.

cause it makes the lamp voltage increase — and hence the re-ignition voltage — and thus ultimately causes the lamp to fail. The luminous efficacy of the new SON lamps therefore decreases much more slowly than that of lamps with mercury alone as the buffer gas, and their life is longer.

Table I. The most important characteristics of three high-pressure sodium lamps of the new generation (*SON**), compared with those of existing lamps (*SON*).

Type	70 W		150 W		400 W	
	SON	*SON**[a]	*SON*	*SON**	*SON*	*SON**[a]
Luminous flux (lm) (clear envelope)	6000	7000	14500	16000	47000	56000
Luminous efficacy (lm/W)	86	100	95	106	120	140
Partial pressures (kPa; ±20%) in the operating lamp:						
sodium vapour	20	20	15	15	12	12
mercury vapour	200	200	150	150	120	120
xenon	25	400	25	250	25	400

[a] experimental unit

Summary. Extensive research into the high-pressure sodium discharge has resulted in the development of a new generation of high-pressure sodium lamps, which have a luminous efficacy about 15% better than that of existing SON lamps and also a longer life, but in other respects differ very little. The improvement is based on the use of xenon as the buffer gas. The xenon is not used on its own, because this would result in a longer and thinner discharge tube, an unsuitable chromaticity and a very high ignition voltage. The first two of these disadvantages have been overcome by adding a certain amount of mercury to the xenon; the third disadvantage has been eliminated by using an auxiliary electrode consisting of a wire stretched along the outside of the discharge tube and connected to one of the main electrodes through a capacitor or a bimetallic strip. A 150 W lamp of the new generation is already in production; other lamps for higher and lower powers are being developed.

Station and programme identification in FM sound broadcasting

G. C. M. Gielis, J. B. H. Peek and J. M. Schmidt

FM sound-broadcast transmissions in the Netherlands are at present accompanied by a coded digital message containing information about the programme and the identity of the station. Similar test transmissions are being made in other Western European countries. An appropriately designed receiver can decode this message and display it in the form of letters and figures. This system, called SPI (for Station and Programme Identification), is a useful tuning aid. Extensions to the system, such as automatic search for the strongest signal carrying a given programme, are being studied, with particular attention to car radio.

Introduction

FM reception at home

Many radio listeners find it difficult to tune in their FM receivers. One of the main reasons for this is that the tuning scales of FM receivers only indicate the frequencies and the channel numbers of the stations. They do not give the names of the stations, as on the older medium- and long-wave radio receivers.

The frequency or the channel number of the particular station required has to be looked up in a magazine or the newspaper, or memorized. In fact, however, it is practically impossible to remember all the channels, because there are so many. In the south of the Netherlands it is possible to receive about 25 different FM stations with good stereo quality. Another problem is that since the stations are very close together in the FM band a small movement of the tuning knob retunes to another station, which might be transmitting a different programme. Very often the same programme is being transmitted by several stations and the listener will want to tune in to the nearest station to obtain the best reception.

Listeners using an FM tuner connected to a community-antenna system (CATV) are sometimes faced with yet another problem: in some CATV systems frequency transformation takes place in the central receiving installation so that the listener has to draw up a table to find the new frequencies of the stations.

Ing. G. C. M. Gielis, Dr Ir J. B. H. Peek and J. M. Schmidt are with Philips Research Laboratories, Eindhoven.

Because of these various difficulties it will be much easier to tune in an FM receiver if it includes a display showing

— the programme being received, and
— the transmitter it originates from.

Fig. 1 illustrates an example of such a receiver. The alphanumeric information is transmitted by the station at the same time as the regular programme. The information takes the form of a digital signal modulating a subcarrier. In this article we shall describe such a system by the term 'Station and Programme Identification' (SPI) [1][2].

The information about the location of the station is important because it enables the listener to tune manually to the closest station (and hence the strongest signal). A single programme-identification code would be sufficient, however, to set up the receiver so that it could make its own search for the strongest signal carrying the particular programme required. The part of the code giving the transmitter location could then be omitted.

If further information of variable content is also transmitted in addition to the alphanumeric SPI information always present (name of programme and location of station), a number of other interesting possibilities arise; the receiver can be programmed to switch on automatically for a particular programme, or to record it on a tape recorder. This can be done by giving each particular radio programme its own code and by setting the receiver, pre-programmed with this

code, on stand-by. The indication *CM6* in fig. 1 is an example. This indicates the sixth piece of classical music on that particular day; this code would be given in the published radio programmes. From the moment that the particular programme begins the code *CM6* is added to the non-changing SPI information and thus switches on the audio amplifier or tape recorder.

The potential applications described above — and they could be extended even further — are for the most part little more than dreams. In the Netherlands, however, the SPI signal is already being transmitted continuously and this is also done in Germany, France, Austria and Switzerland from time to time. The intention is to test the reliability of the transmis-

Fig. 1. The 794 receiver, with a display for SPI messages. The display indicates that the receiver is tuned to the programme Hilversum 4, transmitted by the station at Lopik and that at that moment the particular programme item being transmitted is *CM6* (classical music No. 6). The display is a vacuum fluorescent display (VFD).

The receiver can be programmed in several different ways to receive a given programme item or several items one after another. It would be possible, for example, to use thumbwheels, or a light pen that could read bar codes in a printed programme. These would be like the bar codes that are placed on products in supermarkets and used in library books: a black-and-white striped pattern that represents a given sequence of decimal digits. The code can be read by passing a light pen over the strip pattern. It is also possible for the receiver to be programmed so that it will only come on for certain types of programme such as news, classical music, popular music, sport, etc.

It would also be possible to meet one need that has been expressed by radio listeners for years [3]. This is that there should be automatic adjustment of separate sound levels for speech and music. The listener can then use his volume control to select a sound level for speech independently of the desired music level. If music and speech are separately coded in a programme, the appropriate sound level can be selected automatically in the receiver after decoding.

sion of the SPI data. Wave interference between signals reaching the antenna along different paths (multipath reception) and unwanted signals can lead to errors. This is a particular problem for FM reception in vehicles.

FM reception in vehicles

Drivers who wish to listen to a particular programme on the FM band during a journey are faced with quite a different set of problems from those con-

[1] J. B. H. Peek and J. M. Schmidt, A 'Station Programme Identification' (S.P.I.) system for FM sound broadcasting, Int. Broadcasting Conv., London 1978 (IBC 78; IEE Conf. Publn No. 166), pp. 321-323.

[2] As far as we know, the use of such a system was first proposed by L. J. M. van Boldrik of Philips and H. J. van der Heide of the Netherlands Broadcasting Corporation (NOS). Van der Heide has also contributed towards the localization of subcarrier frequencies at $\frac{7}{8}$ and $\frac{9}{8} \times 19$ kHz.

The European Broadcasting Union (EBU) are also known to have an interest in an identification system for FM radio broadcasting; see G. Plenge, Überlegungen zur Frage der optimalen Nutzung von Zusatzinformationskanälen im UKW-Rundfunk, Rundfunktech. Mitt. **24**, 203-206, 1980.

[3] K. Ilmonen, PI signals, a potential ray of hope for frustrated listeners, E.B.U. Rev., tech. Part, No. 142, pp. 284-287, 1973.

fronting the listener at home. A driver who tunes in to a particular programme at the start of a journey will often find that he has to retune to find another frequency carrying the programme he wants as his position relative to the transmitters changes; this is because of the restricted range of all FM transmitters (50 kilometres or less for FM stereo). A short time ago Philips brought out a car radio, the MCC (microcomputer controlled) receiver, which has a microprocessor and a programmable memory and offers an answer to this annoying problem (see *fig. 2*). This

Fig. 2. The MCC (microcomputer-controlled) car radio has a memory in which ten station frequencies can be stored for each of six different programmes (*P1* to *P6*). The receiver automatically searches for the frequency with the strongest signal.

radio can store 60 FM-station frequencies in its memory in groups of ten. Each group consists of station frequencies carrying the same programme. Each of the programme pushbuttons (*P1* to *P6*) can be programmed for such a group, up to a maximum of ten frequencies.

Before the journey the frequencies of the stations transmitting the desired programme in the region of travel must be indicated by pressing one of the pushbuttons. At the start of the journey the pushbutton for the programme desired is pressed. The receiver automatically searches for the frequency that gives the strongest signal and tunes in to it. If this signal falls below a certain level during the journey, the receiver automatically starts to search again for the frequency giving the strongest signal. In this way the desired programme can be received throughout the journey, and the changes from station to station go almost unnoticed.

It occasionally happens, however, that certain frequencies carry a regional programme during part of the day and the main programme at other times. It could then be possible to receive the wrong programme. However, by pressing a special knob on the MCC receiver, the microprocessor is instructed to ignore the regional programme until further notice.

It is worth while investigating whether permanent tuning of a car radio to a desired programme could also be effected by transmitting identification codes with all the programmes. There would have to be a guarantee of sufficiently reliable reception in varied terrain. To find out how such a coded message would be received in mountainous terrain, where the numerous reflections often given unwanted multipath reception, we made some test runs in Austria. The Austrian authorities cooperated by transmitting SPI information, and we found that this was correctly received everywhere where the radio reception was good, provided error detection was employed. However, it is probably not a good idea for a full SPI message to be shown on a car-radio display panel, as on a domestic radio receiver, because it might distract the driver. A shorter message would be sufficient here.

In this article we shall describe some of the systems and circuit aspects of SPI and other identification systems that we have studied since 1975 in the laboratory and during test transmissions.

Data-transmission rate and repetition frequency of the message

In designing a system it is necessary to select an appropriate repetition frequency for the message and to determine the magnitude of the resultant data-transmission rate. The data-transmission rate is the product of the total number of bits per message and the frequency at which the message is repeated.

It is useful here to make a distinction between a domestic radio receiver and one in a motor vehicle.

Domestic receiver

The message repetition frequency is made high enough to ensure direct read-out of the message during manual tuning. Note that the tuning rate is determined here by the rate at which the successive SPI texts can be read.

In 'search tuning' (scanning the band automatically), on the other hand, a maximum search time of, say, ten seconds must not be exceeded. During this time the receiver scans the band for the desired code

(this could be any desired part of the SPI code). Assuming that a maximum of 25 stations can be received and that the transition time between two tuning frequencies during the scan is negligible compared with the acquisition time for the SPI message (two messages at most) this requirement will be met with a repetition frequency of about five messages per second. The required rate of manual tuning is also obtained at this repetition frequency, since it is impossible to read more than 25 texts in ten seconds.

There are various ways of transmitting the alphanumeric SPI message [1]. We shall restrict ourselves here to the method in which the alphanumeric characters are transmitted by means of a six-bit code. Each separate alphanumeric character is represented by a six-bit code. We believe that all the facilities discussed in the introduction can be provided by about 19 alphanumeric characters. For synchronization an initial 'starting code' of 14 bits is required; the complete message will then consist of $14 + 19 \times 6$ bits = 128 bits. With a message repetition frequency of about five messages per second this will give a data-transmission rate of about $128 \times 5 = 640$ bits/s.

Car radio

As we said in the introduction, anyone listening to a radio in a car would like to be able to receive a particular programme continuously throughout the journey. This facility can be provided in various ways.

Firstly, by using the MCC car radio mentioned earlier; however, this means that the user has to go to the trouble of programming the receiver.

Secondly, by using SPI in conjunction with a second receiver in parallel. This method requires further explanation. It assumes a situation in which all the stations transmit an SPI message in addition to the broadcast programme. The detection of such a message at a repetition frequency of five messages per second requires at least 200 milliseconds. The scan searching of the frequency band, with the receiver returning to its original tuning frequency after each identification, introduces a break of at least 200 ms in the received programme. A break as long as this is annoying, particularly since these breaks do not stop until the receiver has scanned the entire band and has found the frequency with the strongest signal carrying the desired SPI code. Test measurements have shown that breaks lasting between 10 and 30 ms are not annoying but can still be heard. In the MCC car radio the duration of the breaks is in this range. It has also been found empirically that breaks of less than 10 ms are inaudible [4].

The annoying breaks in the audio signals, due to the identification of the various SPI messages by the receiver, can be prevented by making use of a second receiver. This second receiver scans the frequency band continuously, identifies the SPI messages and stores some of the data in a memory, while the first receiver continues to provide the broadcast programme without interruption. One of the disadvantages of this method is that a second receiver adds to the expense.

Then there is a third method. Here a single receiver is used, with a code of short duration. It has already been noted that breaks in the received programme are not annoying provided they are shorter than 30 ms. For the breaks to be shorter than 30 ms the identification code must not occupy more than 20 ms, since it takes about 10 ms to retune the receiver and synchronize the decoding stage to the incoming bit string (the synchronization takes most of the 10 ms). This is equivalent to a minimum repetition frequency of 46 messages per second.

Since the data-transmission rate should not be much higher than 600 bits/s, as we shall explain later, the shortened message may consist of a maximum of 13 bits. This is much less than the 128 bits of the SPI message. However, as we mentioned earlier, someone listening in a vehicle is mainly interested in listening to the selected programme without interruption. To achieve this aim it is not necessary to use an SPI code; it is possible to manage with a code with fewer bits that indicates the programmes in each case and does not necessarily have to be composed of alphanumeric characters. We have developed a 13 bit code that can be used to characterize 632 different programmes; the bits also provide the synchronization. This is more than sufficient to identify all present and future programmes within Europe. Test transmissions have already started.

Finally there is a fourth possible method, in which a transmitter includes in the SPI message the frequencies of other transmitters sending out the same programme. The receiver then has to search for the strongest signal on these frequencies when the field-strength has become too low.

An advantage of the fourth method is that the frequencies can be communicated on the same subcarrier as the SPI message itself. Only one subcarrier is then necessary for a domestic or car radio.

The third and fourth methods require error-free data transmission. Further experiments will be necessary before we can develop methods for meeting this requirement in the high-noise conditions of car-radio reception.

[4] D. J. H. Admiraal, B. L. Cardozo, G. Domburg and J. J. M. Neelen, Annoyance due to modulation noise and drop-outs in magnetic sound recording, Philips tech. Rev. 37, 29-37, 1977.

Selection of subcarrier frequency and compatibility

In Western Europe there are an estimated several hundred million FM receivers in use. When a very useful tuning aid such as SPI is introduced there must of course be no audible interference introduced in these existing receivers. In choosing the frequency of the subcarrier that is modulated by the SPI message this requirement for compatibility must be taken into account.

So that we can choose the frequency of the carrier signal, we consider the frequency spectrum of the FM stereo multiplex signal (*fig. 3*). This is the spectrum of the signal immediately after the FM detector. The multiplex spectrum is formed from the mono signal *M* (left-hand signal plus right-hand signal), which extends to 15 kHz, the stereo signal *S* (left-hand signal minus right-hand signal), which modulates a suppressed subcarrier at 38 kHz and extends from 23 to 53 kHz, and the pilot tone (19 kHz). This pilot tone gives the FM carrier a frequency deviation of ± 7.5 kHz, equal to 10% of the maximum deviation. The receiver derives from this pilot tone a 38 kHz subcarrier signal for use in demodulating the stereo signal.

Depending on the location of the subcarrier in the spectrum and on the receivers, audible interference may arise, e.g. because of incorrectly designed stereo decoders. The subcarrier interferes with multiples of the 19 kHz pilot frequency and this produces audible difference frequencies. Interference can also arise because of nonlinear operation before the stereo decoder, such as multipath reception, inaccurate tuning or non-constant group delay in the passband of the i.f. amplifier.

The multiplex spectrum (the MPX spectrum) has a number of gaps in which a subcarrier can be located: the gaps between 15 and 19 kHz, between 19 and 23 kHz and the gap above 53 kHz, which can extend as far as about 100 kHz, depending on the i.f. bandwidth.

In some countries the spectrum above 53 kHz is used for broadcasting an extra programme. In the United States for example 67 kHz is used as a frequency-modulated subcarrier for a special programme (services for the disabled, background music for department stores). This system is known as 'Subsidiary Communications Authorization' (SCA). In West Germany, Austria and Switzerland a traffic-information service known as ARI ('*Autofahrer-Rundfunk-Information*') is operated for drivers. A traffic-information transmission is identified by a constant tone of 57 kHz transmitted with a frequency deviation of ± 3.75 kHz (5%), and a tone of 125 Hz is transmitted whenever a traffic report starts. Transmission of any one of six tones located between 23.75 Hz and 53.98 Hz can be used to identify six different geographical regions.

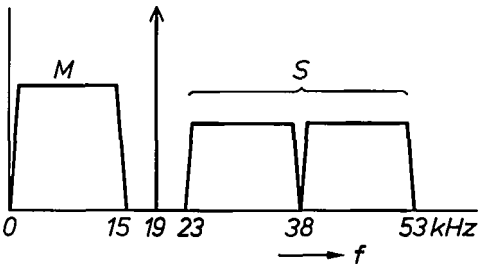


Fig. 3. Frequency spectrum of the FM stereo multiplex signal. *f* frequency. *M* mono signal, the sum of the signals for left-hand and right-hand channels. *S* is the stereo signal, the signal for the left-hand channel minus that for the right-hand channel; it amplitude-modulates a 38 kHz subcarrier; this carrier is suppressed in the transmitter and reconstructed in the receiver from the 19 kHz pilot tone that is also transmitted.

If we wish to have a universally applicable system for station and programme identification then we shall have to take account of the above systems in the band above 53 kHz. To generate a 38 kHz signal, modern radio receivers usually employ an integrated phase-locked loop (*fig. 4*). The phase-locked loop consists of a voltage-controlled oscillator (*VCO*) that produces a square-wave voltage at 76 kHz in the locked state. Halving this frequency in a divider results in a symmetrical square wave at 38 kHz, which only contains odd harmonics. This square-wave voltage is used as the subcarrier signal for demodulating the stereo signal at 38 kHz. Dividing by two again gives a square-wave voltage at 19 kHz. This is used as

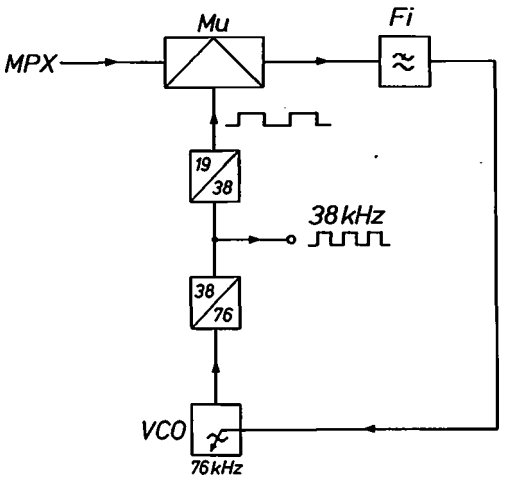


Fig. 4. Phase-locked loop for generating the 38 kHz subcarrier for FM stereo reception. The phase of the subcarrier is controlled by comparing the 19 kHz signal derived from it with the 19 kHz pilot tone present in the received multiplex signal *MPX*. This comparison is made by multiplication in the circuit *Mu*. Depending on the phase relationship this circuit produces a positive or negative direct voltage; when the phase relationship is correct this voltage is zero. The voltage controls the voltage-controlled oscillator *VCO* (after filtering in the lowpass loop filter *Fi*).

the control signal and demodulates the 19 kHz pilot tone in the MPX signal. In the locked state of the loop the two 19 kHz signals will differ in phase by 90°.

Mixing of the 19 kHz square-wave voltage with the MPX signal also results in frequency contributions in the loop from the spectrum around 57 kHz, since the MPX spectrum around 57 kHz will be demodulated by the third harmonic of the 19 kHz square-wave voltage. Even if no additional signals above 53 kHz are present, noise around 57 kHz will be demodulated, which means that the phase of the 38 kHz square wave will begin to fluctuate somewhat, causing distortion in the demodulated stereo signal. However, if there are additional signals with spectrum contributions near to 57 kHz that give rise to demodulation products that get through the loop filter *Fi* (fig. 4), these can also be the cause of audible interference [5].

This is one reason for not locating the SPI subcarrier signal near to 57 kHz, and so we should look more closely at the two gaps in the band around 19 kHz. These gaps also have the advantage that the noise here is about 10 dB less than at 57 kHz, so that a weaker subcarrier with only a third of the frequency deviation may be sufficient.

We have found two frequency bands around 16.625 kHz ($\frac{7}{8}\times 19$ kHz) and 21.375 kHz ($\frac{9}{8}\times 19$ kHz) in which signals do not give audible interference if their level is low enough. 'Low enough' here means that the maximum deviation of the modulated FM carrier is no greater than ± 250 Hz. This subcarrier level is 30 dB below the level recommended by the CCIR for the 19 kHz pilot tone (± 7.5 kHz).

This choice of frequency bands also means that the SPI signals do not cause any increase in interference in an adjacent channel. Laboratory trials with various receivers have shown that if the data-transmission rate for the above two subcarriers does not become substantially greater than 600 bits/s, there will be no audible interference at a maximum deviation of ± 250 Hz.

Practical tests are also being carried out. Since November 1977 uninterrupted test transmissions have been under way on all three FM networks in the Netherlands, with the main aim of finding out whether the compatibility requirement can be satisfied. In these test transmissions, which are being made in collaboration with the Netherlands Broadcasting Corporation (NOS) and the Netherlands Postal and Telecommunications Service (PTT), a subcarrier of 16.625 kHz is being used and SPI information is being transmitted at a data-transmission rate of about 600 bits/s. The bit stream is modulated in two-phase phase-shift keying (PSK) (see *fig. 5*) [6]. The maximum deviation of the main carrier resulting from this SPI signal was originally ± 250 Hz but was increased to ± 500 Hz in November 1978.

Since the SPI signal was added to the transmitted programmes in November 1977 there have been no complaints from listeners. During the Dutch test transmissions no measures were taken initially to limit the audio bandwidth to 15 kHz. This meant that now and again spectral components of the audio signal arose in the SPI channel and sometimes caused errors in the decoded SPI message. These errors were later eliminated by restricting the audio bandwidth to 15 kHz at the transmitter by means of lowpass filters.

Test transmissions have also been made since January 1981 with a shortened message in a 13 bit code on a subcarrier at 21.375 kHz ($=\frac{9}{8}\times 19$ kHz). The frequency deviation caused by each of the two subcarriers has been reduced in this case to ± 250 Hz.

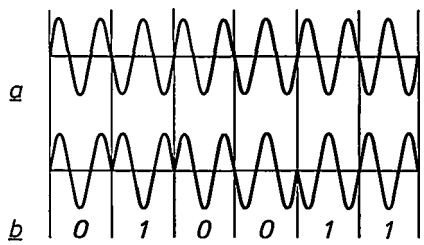


Fig. 5. Two-phase PSK ('phase-shift keying'). *a*) Carrier. *b*) The modulated carrier for the digital signal '010011'.

The SPI demodulator

During the last few years we have been gaining experience with various types of SPI demodulator. A block diagram of an SPI demodulator that has been found to operate very satisfactorily during laboratory trials and test transmissions is shown in *fig. 6*.

The PSK signal and the 19 kHz pilot tone are separated from the received signal by means of a single circuit tuned to 16.625 kHz. The 19 kHz pilot tone is then recovered in a phase-locked loop and the 16.625 kHz SPI signal is mixed with the 19 kHz pilot tone.

The SPI information is thus transferred to a PSK signal at 2.375 kHz. This signal is coherently demodulated in the demodulator with the aid of a 2.375 kHz carrier. This carrier is obtained by dividing the 19 kHz signal. However, its phase still has to be corrected. This is done by passing the signal through a

[5] J. Mielke, Grenzen für die Übertragung von Zusatzinformationen im UKW-Hörrundfunk, Rundfunktech. Mitt. **25**, 74-80, 1981.

[6] More information is given in F. W. de Vrijer, Modulation, Philips tech. Rev. **36**, 305-362, 1976.

phase-shift network controlled by the comparison product of two 4.75 kHz signals.

Signals at twice the frequency are used here because the carrier is not present in the 2.375 kHz PSK signal averaged over a longer time (the signal is as often in phase as it is out of phase; see fig. 5). Squaring makes the sign always positive, and a frequency of 4.75 kHz is thus obtained. A 4.75 kHz carrier is also derived by frequency doubling from the 2.375 kHz carrier required for synchronous detection. Phase comparison of the two 4.75 kHz signals produces a control signal that gives the 2.375 kHz carrier the correct phase. Although the 4.75 kHz signal obtained by squaring contains a great deal of noise, the system described has been shown to operate satisfactorily in practice.

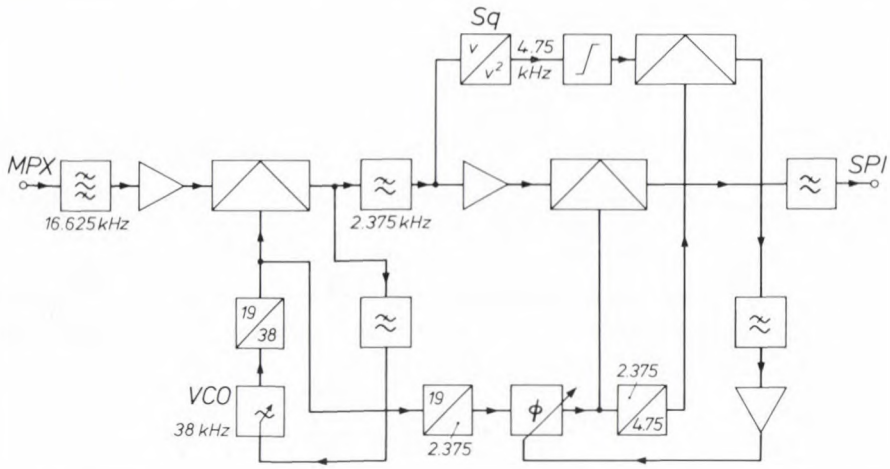


Fig. 6. Block diagram of SPI demodulator. The SPI signal contained in the multiplex signal *MPX*, present in two-phase PSK modulation of a subcarrier at 16.625 kHz, is first transformed into a signal at 2.375 kHz and then demodulated by synchronous detection. The 2.375 kHz square wave required for this is obtained by frequency division from 38 kHz. The correct phase is set in a phase-shift stage ϕ, which is controlled by a voltage obtained by comparing the 2.375 kHz square wave with the signal to be demodulated. This comparison is made at twice the frequency, i.e. 4.75 kHz; the signal to be demodulated is squared in the stage *Sq*.

By linking the data-transmission rate to 19 kHz (a bit rate of 593.75 bits/s is equal to $\frac{1}{32} \times 19$ kHz, for example), very reliable bit synchronization can be obtained.

Error probability

One measure of the quality of an FM receiver is the signal-to-noise ratio $(S/N)_a$ in the audio-frequency band, measured for an audio tone at say 1 kHz. The maximum deviation for the modulation of the tone and the receiver input voltage at which the signal-to-noise ratio is measured must be known.

The signal-to-noise ratio in the SPI channel $(S/N)_{SPI}$ can be expressed as a function of the audio signal-to-noise ratio. The error probability in the SPI signal is in turn a function of $(S/N)_{SPI}$.

The signal-to-noise ratio $(S/N)_{SPI}$ is difficult to measure, but can be calculated from $(S/N)_a$; the error probability in the SPI transmission can then be calculated from established formulae for the error probability in coherent PSK [7].

Let us assume that W represents the audio bandwidth (15 kHz) and $G(f)$ the power spectral-density function of the noise in the audio band after the FM demodulator. This power spectrum has a parabolic shape across the baseband and is expressed by the relation $G(f) \alpha f^2$, where α is a constant. If an RC de-emphasis filter with a 3 dB bandwidth $f_c = 1/2\pi RC$ is used, the total noise power in the audio channel is

$$N_a = \alpha \int_{-W}^{W} \frac{f^2}{1 + \left(\frac{f}{f_c}\right)^2} \, df = 2\alpha f_c^2 \left(W - f_c \arctan \frac{W}{f_c} \right).$$

If the digital SPI information is transferred via a transmission channel that satisfies the Nyquist condition (bandwidth as small as possible, but such that the symbols just do not perturb one another) the baseband width is $B = 1/2T$ Hz where T is the duration of one bit. The total noise power in the modulated SPI channel is

$$N_{SPI} = 2\alpha \int_{f_0-B}^{f_0+B} f^2 df,$$

where f_0 is the subcarrier frequency. This reduces to:

$$N_{SPI} = \alpha(4f_0^2 B + \tfrac{4}{3}B^3) \approx 4\alpha f_0^2 B, \qquad B \ll f_0.$$

From this it follows that the signal-to-noise ratio in the SPI channel is equal to

$$\left(\frac{S}{N}\right)_{SPI} = \left(\frac{S}{N}\right)_a \cdot \frac{P_{SPI}}{P_a} \cdot \frac{N_a}{N_{SPI}}$$

$$= \left(\frac{S}{N}\right)_a \cdot \frac{P_{SPI}}{P_a} \cdot \frac{f_c^2 W - f_c^3 \arctan \frac{W}{f_c}}{2 f_0^2 B},$$

where P_{SPI} is the mean power of the PSK signal and P_a is the mean power of the audio tone.

The error probability can also be measured directly, of course. A comparison of the calculated and measured error probabilities is given in *Table I* for two values of the input voltage to the receiver.

Table I. The error probability P_e in the transmission of the SPI message, calculated from the signal-to-noise ratio $(S/N)_{SPI}$ and measured.

Input voltage of receiver	$(S/N)_{SPI}$	P_e	
		Calculated	Measured
4 μV	2.27 dB	3×10^{-2}	6×10^{-2}
6 μV	5.79 dB	3×10^{-3}	5×10^{-3}

The error probability has been measured on a Philips RH 741 receiver for an SPI bit stream of 600 bits/s and a subcarrier at 16.625 kHz with a maximum deviation of ± 250 Hz. The calculated value of $(S/N)_{SPI}$ is based on a measured audio signal-to-noise ratio $(S/N)_a$ of 43.2 dB for a 1000 Hz tone, modulated with a maximum deviation of ± 22.5 kHz, with a voltage of 4 μV across 60 Ω at the input of the receiver.

These results show that there is reasonable agreement between the calculated and measured error probabilities. In practice we have found that the alphanumeric information can be reliably reproduced with an error probability of 5×10^{-3}, as found for an input voltage of about 6 μV.

Multipath reception

The nonlinear distortion of the audio signal that occurs because the FM signal arrives in the receiver along different paths as a result of reflections (multipath reception) was discovered early in the history of FM broadcasting [8]. This distortion, which can also cause problems in the transfer of the SPI message, is observed mainly in mountainous and hilly regions and in populated areas with high buildings. Where there is multipath reception, the audio signals may also be perturbed as a result of interference with the SPI subcarrier. These effects, which occur for an SPI subcarrier in the vicinity of 19 kHz, are generally very minor because of the small deviation acquired by the carrier. However, multipath reception has a considerable effect on SPI transmission, for reasons that we shall explain later.

In New York and Tokyo the echo patterns that occur during FM broadcast transmissions have been measured. In both cases the carrier frequency was about 450 MHz. In New York (transmitter and receiver both *in* the city) it was found that the echo pattern varied considerably over short distances and the echo delay was not more than 10 μs [9]. In Tokyo (receiver in the mountains *outside* the city) delays of up to 150 μs were recorded and echo levels of −10 to −25 dB as compared with the direct signal [10].

These measurements all relate to stationary reception; in mobile reception, as with a car radio, allowance has to be made for rapid variation in the echo pattern, with the signal regularly fading and reappearing [11]. In addition, measurements have also been made of the subjective annoyance in two-path reception (a single echo) of FM broadcast transmissions [12]. The conclusion arrived at here was that the subjective annoyance increases with the delay and the level of the audio signal (i.e. with the deviation).

The nonlinear distortion that occurs as a result of stationary multipath reception is difficult to calculate even in the case of ideal amplitude-limiting before demodulation and with simple modulating signals. Because of this difficulty and the complexity of the MPX signal we have constructed an echo simulator that can simulate a single echo. With a delay line delays of 10 μs to 70 μs can be introduced in steps of 10 μs.

We have made distortion measurements with this echo simulator. For the audio signal we used white noise up to a cut-off frequency of 15 kHz, so that the MPX spectrum consisted of noise with a power spectrum that was flat up to 15 kHz and between 23 kHz and 53 kHz, and a pilot tone at 19 kHz. *Fig. 7a* clearly shows the various gaps in the MPX spectrum. Fig. 7*b* shows the MPX spectrum after FM demodulation, when there is a single echo with a delay of 70 μs and a level of −12 dB with respect to the main signal. It can clearly be seen that the intermodulation noise generated has now filled up the gaps and can perturb an identification signal in these gaps. It can also be seen that the level of the intermodulation noise spectrum is higher in the neighbourhood of 19 kHz than immediately above 53 kHz.

Measurements of the signal-to-intermodulation-noise ratio in the SPI signal at 16.625 kHz lead to the conclusion that for a given echo level the signal-to-intermodulation-noise ratio has a tendency to decrease if the echo delay increases, and that when there is an increasing level of the modulating noise signal (higher deviation), the signal-to-intermodulation-noise ratio decreases.

[7] M. Schwartz, W. R. Bennett and S. Stein, Communication systems and techniques, McGraw-Hill, New York 1966.
[8] M. S. Corrington, Frequency-modulation distortion caused by multipath transmission, Proc. I.R.E. 33, 878-891, 1945.
[9] W. R. Young, Jr., and L. Y. Lacy, Echoes in transmission at 450 megacycles from land-to-car radio units, Proc. I.R.E. 38, 255-258, 1950.
[10] CCIR Doc. II/22-E (1962, Japan): Distortion in frequency modulation receivers due to multipath propagation.
[11] W. C. Jakes, Jr. (ed.), Microwave mobile communications, Wiley, New York 1974.
See also: R. C. French and P. J. Mabey, Error control in mobile-radio data communication, Philips tech. Rev. 39, 172-182, 1980 (No. 6/7).
[12] Th. Bossert, Beurteilung der Qualität des UKW-Empfanges hinsichtlich der Störungen durch Mehrwegausbreitung, NTG-Fachberichte 72, 227-235, 1980.

It follows that in situations where the distortion of the audio signal increases or decreases, the perturbation of the SPI signal will also increase or decrease. There is, however, one important case in which the SPI signal remains unperturbed in multipath reception, regardless of the number of echoes and their relative levels. This situation arises during the pauses in music or speech. During these pauses the multiplex signal consists only of the 19 kHz pilot tone plus the weak SPI subcarrier. The phase modulation of the FM signal during multipath reception is a periodic function with 19 kHz as the fundamental frequency, so that the demodulated spectrum will consist of a 19 kHz fundamental and its harmonics (38 kHz, 57 kHz, etc.) regardless of the echo pattern. This information can be put to use in designing an error-detection circuit.

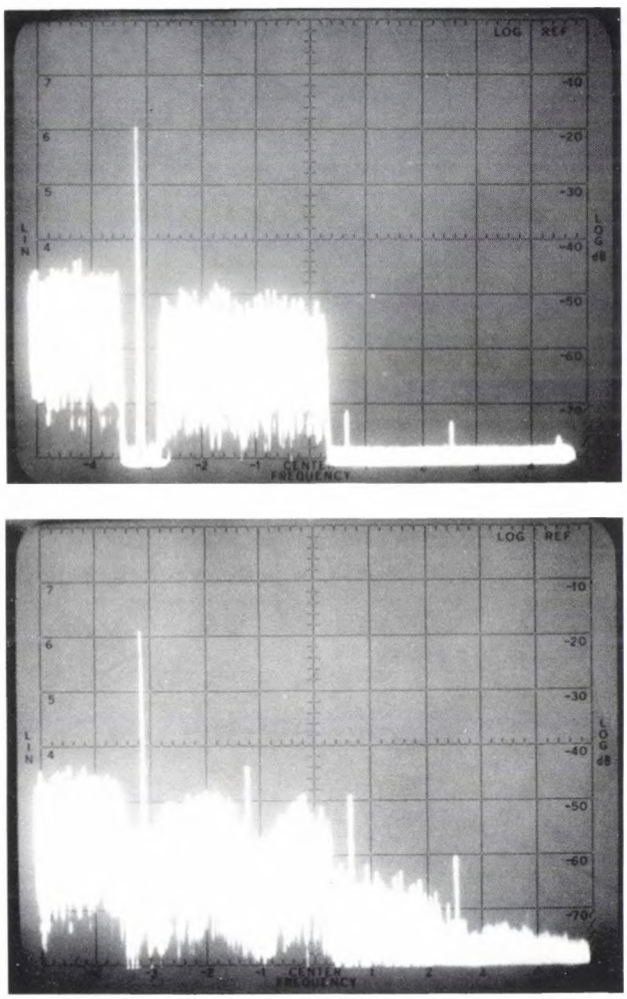


Fig. 7. FM-MPX spectrum when the transmitter is modulated by two independent frequency bands of white noise 15 kHz wide. *a*) No echo. The mono signal from 0 to 15 kHz, the pilot tone at 19 kHz and the stereo signal from 23 to 53 kHz can all be seen in the spectrum. There is a gap on either side of 19 kHz and above 53 kHz. *b*) Besides the direct signal an echo, 12 dB weaker, is received after 70 μs. The gaps in this spectrum have been filled.

The design approach is based on the continuous repetition of the identification messages. If for example the bits of a character are not the same in three successive messages, then this indicates a transmission error. If on the other hand the bits are always the same, this indicates that the character has been received correctly. A bit pattern that has been examined for errors in this way and found to be correct can then be stored in a random-access memory (RAM). The information stored in the memory is displayed on the display panel. Only when another bit pattern is found in which the bit patterns are the same in three successive messages is new information written to the memory. This method has the advantage that no extra coding bits are required, so that for a given transmission rate neither the message repetition frequency nor the number of characters per message need be reduced.

The method would be doomed to failure in the cases in which the signal-to-noise ratio in the SPI channel was poor over a long period. We have seen, however, that during pauses in the radio signal there is no intermodulation distortion in the SPI channel, so that error-free reception of the SPI message is guaranteed during these pauses. And, fortunately, a reasonable percentage of the time occupied by speech and music consists of pauses.

To prevent the system from being delayed by the comparison of three successive messages, which can be somewhat undesirable, particularly during tuning, the first available message is written directly to a still empty memory. If there is a good signal-to-noise or signal-to-interference ratio the message displayed immediately will be practically free of error. If signal conditions are poor, on the other hand, there will be some delay before the correct message appears on the display.

Fig. 8 is a block diagram of a system arranged in this way. A precondition for decoding the SPI bits is the recovery of the clock signal. This can be achieved by means of a phase-locked loop. If the SPI signal is perturbed by multipath reception the clock signal may be lost temporarily; processing of the information already stored in the memory is then impossible.

To solve this problem we have brought the clock frequency into a simple numerical ratio to the 19 kHz pilot frequency, by making its frequency $\frac{1}{32} \times 19$ kHz $= 593.75$ Hz. The 19 kHz pilot tone is a fairly strong signal (10% of the maximum frequency deviation) and much less sensitive to interference than the SPI signal ($\frac{1}{3}$% of the maximum deviation), so that it is less probable that the clock frequency will be per-

turbed in this way. A special circuit has been added to maintain the frequency and also the correct phase of the clock signal during any interference.

The block *ErrDet* in fig. 8 contains an error-detection circuit based on the above principle. The messages each contain 128 bits. The demodulated bit stream is delayed in a shift register of length $2 \times 128 + 4 = 260$ bits. The shift register has take-off points after the sections 1, 2, 3, 4; 129, 130, 131, 132; and 257, 258, 259, 260. This gives three blocks each with four take-off points. If these three blocks are the same then the first bit of the block is written to a memory.

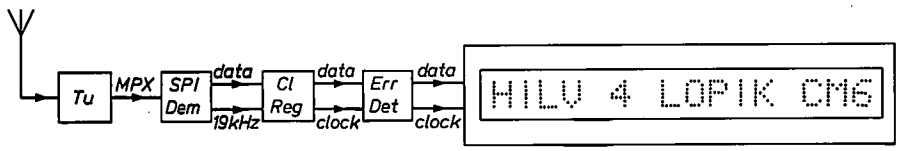


Fig. 8. Processing SPI information in an FM receiver. The FM-MPX signal is available at the output of the tuner unit *Tu*. The SPI demodulator *SPIDem* extracts the binary SPI data and the 19 kHz pilot tone. The clock signal (*ClReg*) is derived from the pilot tone. Data and clock signals pass to a circuit *ErrDet* in which errors are detected by comparing three successive messages. The corrected data is passed to the display.

The advantages to be gained by using the above system under conditions of multipath reception cannot be demonstrated quantitatively. The reason for this is that the intermodulation interference that arises in the SPI channel depends very much on the type of audio signal (speech, classical music, popular music, etc.), the variations in its level and the distribution of the pauses and their duration.

The system described has however been tested with the echo simulator and very varied audio signals have been measured over a long period of time. It has been shown that even when there is multipath reception that distorts the audio signal to an unacceptable extent, the SPI message on the display is unaffected.

The error-detection system described here also improves the sensitivity. It has already been pointed out that without the error-detection system reliable reproduction of the messages is obtained at an input voltage of 6 μV. When the error-detection system is used reliable reproduction can be obtained at an input voltage as low as 2 μV; this corresponds to an increase in sensitivity of almost 10 dB.

Summary. In various Western European countries test transmissions are being made in which FM radio broadcasts are accompanied by a digital 'SPI' message announcing the identity of the station and giving information about the programme. After decoding in the receiver the information can be displayed on an alphanumeric display; this can make tuning easier. The digital signal modulates a weak subcarrier at 16.625 kHz in two-phase PSK (contribution to the frequency deviation is ± 250 Hz). Reliable SPI reception is obtained with a voltage of 6 μV at the receiver input; if errors are detected by comparing successive SPI messages, however, 2 μV is sufficient. Distortion of the message due to multipath reception can also be corrected in this way. A shorter code is being studied for automatically searching for the strongest station transmitting a particular programme; this is of particular interest for car radios.

Scientific publications

These publications are contributed by staff of laboratories and plants which form part of or cooperate with enterprises of the Philips group of companies, particularly by staff of the following research laboratories:

Philips Research Laboratories, Eindhoven, The Netherlands	*E*
Philips Research Laboratories, Redhill, Surrey RH1 5HA, England	*R*
Laboratoires d'Electronique et de Physique Appliquée, 3 avenue Descartes, 94450 Limeil-Brévannes, France	*L*
Philips GmbH Forschungslaboratorium Aachen, Weißhausstraße, 51 Aachen, Germany	*A*
Philips GmbH Forschungslaboratorium Hamburg, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany	*H*
Philips Research Laboratory Brussels, 2 avenue Van Becelaere, 1170 Brussels (Boitsfort), Belgium	*B*
Philips Laboratories, N.A.P.C., 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A.	*N*

G. A. Acket & H. Koelmans: Recent developments in semiconductor injection lasers for optical communication.
Acta Electronica **22**, 295-300, 1979 (No. 4). *E*

J. P. Arragon: Sources émettrices et modulation des diodes électroluminescentes pour transmissions numériques.
Acta Electronica **22**, 323-328, 1979 (No. 4). *L*

J. P. Arragon: Récepteur à photodiode PIN pour transmissions numériques par fibres optiques.
Acta Electronica **22**, 329-334, 1979 (No. 4). *L*

P. W. J. M. Boumans, F. J. de Boer, A. W. Witmer & M. Bosveld: Outline of a method for spectrographic general survey analysis using liquid sampling and an inductively coupled plasma.
Spectrochim. Acta **33B**, 535-544, 1978 (No. 8). *E*

J. C. Brice & A. M. Cole: The characterization of synthetic quartz by using infra-red absorption.
Proc. 32nd Ann. Symp. on Frequency Control 1978, Atlantic City, pp. 1-10. *R*

H. H. Brongersma, C. M. G. Jochem, T. P. M. Meeuwsen, P. J. W. Severin & G. A. C. M. Spierings: The preparation of alkali-germanosilicate optical fibres using the double crucible system.
Acta Electronica **22**, 245-254, 1979 (No. 3). *E*

K. H. J. Buschow: Change in magnetic properties of rare earth-transition metal compounds upon H_2-absorption.
Hydrides for energy storage, ed. A. F. Andresen & A. J. Maeland, pp. 273-285; Pergamon Press, Oxford 1978. *E*

K. H. J. Buschow & A. R. Miedema: Hydrogen absorption in rare earth intermetallic compounds.
Hydrides for energy storage, ed. A. F. Andresen & A. J. Maeland, pp. 235-249; Pergamon Press, Oxford 1978. *E*

J. P. Cabanié: Etude d'un coupleur en étoile pour un réseau de transmissions de données à haut débit.
Acta Electronica **22**, 351-358, 1979 (No. 4). *L*

A. Charles-Georges & R. Gousseau: Etude et réalisation mécaniques d'un microconnecteur pour fibre optique.
Acta Electronica **22**, 335-341, 1979 (No. 4). *L*

S. R. Chinn (M.I.T., Lexington, Mass.) **& W. K. Zwicker:** Thermal conductivity and specific heat of NdP_5O_{14}.
J. appl. Phys. **49**, 5892-5895, 1978 (No. 12). *N*

P. A. Devijver: A note on ties in voting with the *k-NN* rule.
Pattern Recognition **10**, 297-298, 1978 (No. 4). *B*

J. Donjon: Le tube Titus: application à la projection d'images en couleur sur grand écran.
Onde électr. **58**, 558-564, 1978 (No. 8/9). *L*

P. W. East & K. A. White (MEL, Crawley, Sussex): Complexity/performance trade-offs in IFM receiver design.
Conf. Proc. Military Microwaves, London 1978, pp. 16-25. *R*

L. H. M. Engel: Optische telecommunicatie.
Natuur en Techniek **46**, 412-427, 1978 (No. 6). *E*

P. G. van Engen: Mode degeneracy in magnetic garnet optical waveguides with high Faraday rotation.
J. appl. Phys. **49**, 4660-4662, 1978 (No. 9). *E*

A. A. J. Franken & W. P. Weijland (Philips Elcoma Division, Eindhoven): Performance of the 2/3" 'Plumbicon' camera tube and some alternatives.
Int. Broadcasting Conv., London 1978 (IBC 78; IEE Conf. Publn No. 166), pp. 59-62.

R. C. French: Measuring data demodulator noise performance.
Electronic Engng. **50**, Oct. 1978, pp. 33, 36 & 37 (No. 613). *R*

P. Geittner, D. Küppers, H. Lydtin & J. Ungelenk: Optical fibres prepared by the plasma activated chemical vapour deposition method.
Acta Electronica **22**, 237-244, 1979 (No. 3). *A*

R. Genève: Introduction: optical fibres and communications.
Acta Electronica **22**, 179-191, 1979 (No. 3). *(Also in French.)* *L*

W. Goedbloed (Philips Elcoma Division, Eindhoven), **H. Hieber & A. G. van Nie:** Ageing tests on microwave integrated circuits.
Radio and electronic Engr. **48**, 13-22, 1978 (No. 1/2). *H, E*

S. Gourrier, A. Mircea & M. Bacal (Ecole Polytechnique, Palaiseau): Use of multipole plasma for the oxidation of semiconductors.
Gordon Conf., Andover (USA) 1978, 13 pp. *L*

G. J. van Gurp, W. F. van der Weg & D. Sigurd (Research Institute for Physics, Stockholm): Interactions in the Co/Si thin-film system, II. Diffusion-marker experiments.
J. appl. Phys. **49**, 4011-4020, 1978 (No. 7). *E*

J. P. Hazan: Characterization of multimode optical fibre transmission capacity.
Acta Electronica **22**, 203-224, 1979 (No. 3). *L*

J. C. M. Henning & J. H. den Boef: Strain-modulated electron spin resonance of cubic Cr^{3+} in MgO.
Phys. Rev. B **18**, 60-68, 1978 (No. 1). *E*

S. van Heusden & L. G. J. Mans (Eindhoven University of Technology): Alternating measurement of ambient and cabin ozone concentrations in commercial jet aircraft.
Aviation, Space, and env. Med. **49**, 1056-1061, 1978 (No. 9). *E*

L. Jacomme: Quelques aspects théoriques de la propagation dans les fibres optiques multimodes.
Acta Electronica **22**, 193-202, 1979 (No. 3). *L*

M. Janta-Polczyński: Coming to grips with the semantics of natural language: some achievements and some proposals in artificial intelligence.
Studia Semiotyczne VIII, Assolineum, Warsaw-Wroclaw, pp. 91-106, 1978. *B*

C. M. G. Jochem, T. P. M. Meeuwsen, F. Meyer, P. J. W. Severin & G. A. C. M. Spierings: Technology of alkali germanosilicate graded-index fibres.
4th Eur. Conf. on Optical communication, Genova 1978, pp. 2-10. *E*

H. D. Jonker: LPE growth reproducibility of garnet bubble materials.
Mat. Res. Bull. **13**, 921-930, 1978 (No. 9). *E*

G. D. Khoe & G. Kuyt: On the realistic efficiency of coupling light from GaAs laser diodes into parabolic index optical fibres.
4th Eur. Conf. on Optical communication, Genova 1978, pp. 309-312. *E*

W. Kindler, G. Wüster (both with Rhein.-Westf. T.H. Aachen) **& H. Rau:** Equation of state for the vapour of concentrated and diluted hydrochloric acid.
Ber. Bunsen-Ges. Phys. Chemie **82**, 543-545, 1978 (No. 5). *A*

J. T. Klomp & R. H. Lindenhovius: Microstructural and physical properties of Al_2O_3-Fe cermets.
Ceramurgia Int. **4**, 59-65, 1978 (No. 2). *E*

W. L. Konijnendijk & J. M. Stevels (Eindhoven University of Technology): Structure of borate and borosilicate glasses by Raman spectroscopy.
Borate glasses: structure, properties, applications, ed. L. D. Pye, V. D. Fréchette & N. J. Kreidl, pp. 259-279; Plenum Press, New York 1978. *E*

G. Kowalski: Suppression of ring artefacts in CT fan-beam scanners.
IEEE Trans. **NS-25**, 1111-1116, 1978 (No. 5). *H*

D. J. Kroon: Analysis of ambient air.
J. Physics E **11**, 497-507, 1978 (No. 6). *E*

D. Küppers, J. Koenings & H. Wilson: Deposition of fluorine-doped silica layers from a $SiCl_4/SiF_4/O_2$ gas mixture by the plasma-CVD method.
J. Electrochem. Soc. **125**, 1298-1302, 1978 (No. 8). *A*

G. M. Loiacono, W. N. Osborne, M. Delfino & G. Kostecky: Single crystal growth and properties of deuterated triglycine fluoroberyllate.
J. Crystal Growth **46**, 105-111, 1979 (No. 1). *N*

H. Lydtin & F. Meyer: Review of techniques applied in optical fibre preparation.
Acta Electronica **22**, 225-235, 1979 (No. 3). *A, E*

H. H. van Mal & A. R. Miedema: Some applications of $LaNi_5$-type hydrides.
Hydrides for energy storage, ed. A. F. Andresen & A. J. Maeland, pp. 251-260; Pergamon Press, Oxford 1978. *E*

B. J. Mulder & J. J. Vrakking: Beryllium thin films as optical filters in helium discharge lamps.
J. Physics E **11**, 743-744, 1978 (No. 8). *E*

H. J. M. Otten (Philips Elcoma Division, Eindhoven): Technical and commercial aspects of fibre optic components.
Acta Electronica **22**, 285-293, 1979 (No. 4). *(Also in French.)*

J. B. H. Peek & J. M. Schmidt: A 'Station Programme Identification' (S.P.I.) system for FM sound broadcasting.
Int. Broadcasting Conv., London 1978 (IBC 78; IEE Conf. Publn No. 166), pp. 321-323. *E*

J. G. J. Peelen & J. W. Versluis (Philips Glass Development Centre, Eindhoven): Pilot plant production of optical fibres.
Acta Electronica **22**, 255-260, 1979 (No. 3).

J. G. J. Peelen, J. W. Versluis & A. P. Vervaart (Philips Glass Development Centre, Eindhoven): Optical quality of fibers produced with the PCVD process in pilot plant conditions.
4th Eur. Conf. on Optical communication, Genova 1978, Suppl. pp. 69-72.

J. R. Périlhou: L'imagerie médicale par ultrasons et ses limitations.
Ann. Radiol. **21**, 547-550, 1978 (No. 7). *L*

A. Pirotte: Linguistic aspects of high-level relational languages.
Date base technology (Infotech), Vol. 2, pp. 271-300, 1978. *B*

R. J. van de Plassche: A sigma-delta modulator as an A/D converter.
IEEE Trans. **CAS-25**, 510-514, 1978 (No. 7). *E*

M. H. J. van Rijswick: Metal hydride electrodes for electrochemical energy storage.
Hydrides for energy storage, ed. A. F. Andresen & A. J. Maeland, pp. 261-271; Pergamon Press, Oxford 1978. *E*

P. Röschmann, W. Tolksdorf & F. Welz: Annealing effects on cation distribution in diamagnetically substituted single crystal yttrium iron garnet.
IEEE Trans. **MAG-14**, 704-706, 1978 (No. 5). *H*

T. E. Rozzi, J. H. C. van Heuven & G. H. in 't Veld: A new d.h. laser configuration with passive transverse field confinement.
4th Eur. Conf. on Optical communication, Genova 1978, pp. 364-372. *E*

Ch. Sauer, W. Zinn (both with Kernforschungsanlage Jülich) **& W. Tolksdorf:** Mössbauer-effect measurements of transferred hyperfine fields in Al-substituted yttrium iron garnets.
IEEE Trans. **MAG-14**, 701-703, 1978 (No. 5). *H*

H. Schomberg: An improved approach to reconstructive ultrasound tomography.
J. Physics D **11**, L 181-185, 1978 (No. 15). *H*

M. F. H. Schuurmans, D. Polder & Q. H. F. Vrehen: Superfluorescence: QM derivation of Maxwell-Bloch description with fluctuating field source.
J. Opt. Soc. Amer. **68**, 699-700, 1978 (No. 5). *E*

G. F. W. Searle (University of Warwick, Coventry) **& J. S. C. Wessels:** Role of β-carotene in the reaction centres of Photosystems I and II of spinach chloroplasts prepared in non-polar solvents.
Biochim. biophys. Acta **504**, 84-99, 1978 (No. 1). *E*

A. Thayse: Meet and join derivatives and their use in switching theory.
IEEE Trans. C-27, 713-720, 1978 (No. 8). *B*

M. J. Underhill, P. A. Jordan, M. A. G. Clark & R. I. H. Scott: A general purpose LSI frequency synthesizer system.
Proc. 32nd Ann. Symp. on Frequency Control 1978, Atlantic City, pp. 365-372. *R*

J. D. B. Veldkamp & R. J. Klein Wassink (Philips Centre for Technology, Eindhoven): Het nuttig gebruik van breuk bij het verwerken en bewerken van brosse materialen.
Polytechn. T. Procestechniek **33**, 302-308, 1978 (No. 6). *E*

C. H. F. Velzel: Nonlinear processing of the interferogram in Fourier spectroscopy.
J. Opt. Soc. Amer. **68**, 915-919, 1978 (No. 7). *E*

Q. H. F. Vrehen: Superfluorescentie van cesiumatomen.
Ned. T. Natuurk. A **44**, 76-79, 1978 (No. 2). *E*

Q. H. F. Vrehen & H. M. Gibbs (Bell Laboratories, Murray Hill, N.J.): Superfluorescence experiments: a review.
J. Opt. Soc. Amer. **68**, 699, 1978 (No. 5). *E*

L. Vriens: Multistep ionization in the positive column of low-pressure Na-Ne and Ne discharges.
J. appl. Phys. **49**, 3814-3820, 1978 (No. 7). *E*

L. Vriens, R. A. J. Keijser & F. A. S. Ligthart: Ionization processes in the positive column of the low-pressure Hg-Ar discharge.
J. appl. Phys. **49**, 3807-3813, 1978 (No. 7). *E*

G. F. Weston: Alphanumeric display.
Proc. IEE **125**, 1077-1099, 1978 (No. 11R). *R*

Published in ESSCIRC 78, Dig. tech. Papers 4th Eur. Solid State Circuits Conf., Amsterdam 1978:

J. O. Voorman: Design and applications of adaptive gyrators (pp. 15-18). *E*
K. Mouthaan: Trends in optical communication systems (pp. 22-23). *E*
J. Lohstroh: ISL, a fast and dense low-power logic, made in a standard Schottky process (pp. 39-42). *E*
J. Lohstroh: Noise margins of I^2L, SI^2L, ISL and STL (pp. 51-54). *E*
L. Nederlof, H. J. M. Veendrick & A. T. van Zanten: Content addressable memory with parallel write facilities (pp. 92-94). *E*
R. J. van de Plassche & D. Goedhart: A monolithic 14-bit D/A converter (pp. 110-112). *E*
J. P. L. Lagerberg: Vertical yield: an aid in characterising the process yield of integrated circuits (pp. 167-168). *E*
N. F. Benschop & L. C. M. Pfennings: A pipelined array product accumulator in dynamic NMOS for efficient signal processing (pp. 188-190). *E*

Published in Physics of magnetic garnets, Proc. Int. School of Physics 'Enrico Fermi' Course LXX, ed. A. Paoletti, Società Italiana di Fisica, Bologna 1978:

P. Hansen: Magnetic anisotropy and magnetostriction in garnets (pp. 56-133). *H*
U. Enz: Irreversible photomagnetic effects in garnets (pp. 364-378). *E*
R. Metselaar (Eindhoven University of Technology) **& P. K. Larsen:** Electrical properties of yttrium iron garnet (pp. 417-444). *E*
G. B. Scott: The optical absorption and magneto-optic spectra of $Y_3Fe_5O_{12}$ (pp. 445-466). *R*
W. Tolksdorf: Preparation and imperfections of magnetic materials with garnet structure (pp. 521-542). *H*

Supplement to Philips Technical Review
Beilage der Philips Technischen Rundschau
Bijlage van Philips Technisch Tijdschrift

PHILIPS

July 1981

Recent United States Patents

Abstracts from patents that describe inventions from the following research laboratories that form part of or cooperate with the Philips group of companies:

Philips Research Laboratories, Eindhoven, The Netherlands	*E*
Philips Research Laboratories, Redhill, Surrey RH1 5HA, England	*R*
Laboratoires d'Electronique et de Physique Appliquée, 3 avenue Descartes, 94450 Limeil-Brévannes, France	*L*
Philips GmbH Forschungslaboratorium Aachen, Weißhausstraße, 51 Aachen, Germany	*A*
Philips GmbH Forschungslaboratorium Hamburg, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany	*H*
Philips Research Laboratory Brussels, 2 avenue Van Becelaere, 1170 Brussels (Boitsfort), Belgium	*B*
Philips Laboratories, N.A.P.C., 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, USA	*N*

4 222 159

Method of manufacturing a color display tube shadow mask

J. Koorneef *E*

A method of manufacturing a color display tube of the refocusing type in which supports of insulation material are secured against an apertured metal plate. The supports are provided with a conductor at least on the side remote from the plate, so that the plate constitutes a first set of lens electrodes and the conductors constitute a second set of lens electrodes. The two sets of lens electrodes form a quadrupole lens in each aperture in the metal plate when a voltage difference is applied between the first set and the second set. The defocusing direction of the quadrupole lens is parallel to the phosphor strips of the display screen.

---

4 227 532

Device for crushing calculi in the urinary bladder

K. H. Bluhm *H*
H. Kunath

A device for crushing calculi in the urinary bladder, comprising a bundle of elongate, flexible lithotriptors which are accommodated in a ureter catheter and on whereto reciprocating movements can be imparted by a drive system.

---

4 228 505

Method for computed tomography

W. Wagner *H*

In transversal computer tomography apparatus, in which the positioning zone in which the patient can be positioned is larger than the scanning zone in which a body slice can be scanned, reconstruction errors are liable to occur. These errors are caused by incomplete irradiation of the body during examination. They become manifest not only as an incorrect image of the area not irradiated, but also have an adverse effect on the image of the other, completely irradiated areas. The invention enables reduction of these errors.

4 229 682

Electronically commutating motor

B. H. A. Goddijn *E*

An electronically commutating motor having a phase winding connected in at least one of the branches of a bridge circuit. A comparator measures the voltage across one of the diagonals of the bridge and switches the voltage across the other bridge diagonal as a function of the polarity of the first-mentioned voltage.

---

4 229 802

Digital adding device

L. D. J. Eggermont *E*

Digital adding device for determining the sum of a plurality of binary coded numbers, having a digital parallel-accumulator having a first store for storing intermediate sums and intermediate carries which can be taken over into a second store by means of a switching circuit in order to be processed to the final sum in a final adder. This enables the use of the digital parallel-accumulator for determining consecutive sums in a more efficient manner.

---

4 229 805

Magnetic bubble-domain device

D. J. Breed *E*

A magnetic domain device comprising a layer of a magnetic material for the formation of magnetic domains, for example bubbles, under the influence of a bias magnetic field and a propagation structure with magnetically operating elements for driving the magnetic domains by the sequential formation thereon of preferred positions for the domains. The current conductors used to propagate the domains, either meander conductors or rotary field 'coils', are arranged on a layer of silicon, so that the heat developed in the current conductors can readily be carried off by the silicon which exhibits a good thermal conductivity.

4 229 821

System for data transmission by means of an angle-modulated carrier of constant amplitude

F. de Jager E
C. B. Dekker
D. Muilwijk

A system having a transmitter and a receiver for transmitting binary data signals of a given symbol rate, an angle-modulated carrier signal of a substantially constant amplitude and a continuous phase being generated in the modulation stage of the transmitter, and the transmitted modulated signal being orthogonally, coherently demodulated in the receiver. The modulation stage of the transmitter is arranged so that the phase of the modulated signal changes in each symbol interval by an amount from the sequence $-\pi/2$, $-\pi/4$, 0, $\pi/4$, $\pi/2$ (rad), which amount is determined for the relevant symbol interval by at least two successive data symbols, and the value of the phase within the relevant symbol interval is determined by a filtered version of at least these two successive data symbols. These measures result in a system which, without sacrificing the remaining desired communication properties of FFSK-systems, utilizes the available frequency spectrum in a more efficient manner than FFSK-systems, because the modulated signal has both a narrower spectral main lobe and, for frequencies outside this spectral main lobe, considerably less power than the FFSK-signal. Consequently this system is very well suited for efficient data transmission over radio links.

---

4 230 788

Method of manufacturing an external electrically conducting metal pattern.

E. J. Spiertz E
C. F. W. Flinsenberg
L. K. H. van Beek

A method of making an electrically conducting metal pattern on a superficially non-conducting support which comprises imagewise exposing to light a photosensitive material containing either a diazosulfide or a diazosulfonate, which produces a light reaction product which is capable of forming free silver and mercury metal from water-soluble silver and mercurous compounds. A mixture of water and at least one solvent from the group consisting of chloroform, toluene, ethylacetate, liquid alcohols and ketones is used in the treatment of the exposed photosensitive material to form a latent image.

---

4 230 915

Record carrier with an optically readable radiation-reflecting information structure

J. G. Dil E
B. A. J. Jacobs

A record carrier is described having an optically readable radiation-reflecting information structure, which comprises trackwise arranged information areas which, in the track direction and transverse to the track direction, are spaced from each other by intermediate areas. It is demonstrated that if the angle of inclination between the walls of the first areas and the normal to the record carrier has one value between 30° and 65° for a satisfactorily reproducible record carrier, the geometrical distance between the plane of the information areas and the plane of the intermediate areas should have one value between $(165/N)$ nanometers and $(270/N)$ nanometers, N being the refractive index of a transparent medium which is disposed between the first and the second plane.

4 230 939

Information-recording element having a dye-containing auxiliary layer

M. R. J. de Bont E
P. J. Kivits
C. J. Schoot
P. Zalm

The invention provides an information-recording element in which information can be written and read optically. The element is constructed from a transparent substrate, a laser light-absorbing, dye-containing auxiliary layer provided thereon, as well as a laser light-reflecting recording layer present on the auxiliary layer. Upon recording information the element is exposed to pulsated laser light via the substrate, holes being formed in the recording layer. The auxiliary layer stimulates the formation of holes, a saving of laser light energy being obtained. In a favorable embodiment the auxiliary layer has a laser light absorption of from 20 to 80% and a maximum thickness of 250 nm. In a further favorable embodiment the auxiliary layer also comprises an endothermal material, for example, nitrocellulose having a nitrogen content of at least 11%.

---

4 231 035

Liquid crystal display for large time multiplexing factors

C. Z. van Doorn E
J. J. M. J. de Klerk

A liquid crystal display screen of the twisted nematic type comprising a liquid crystal having a dielectric relaxation and a high dispersion of $\varepsilon_{||}$ is used in a display device with a dual-frequency control. This structure enables the implementation of matrix-multiplex display devices having large pluralities of lines which must be successively scanned.

---

4 231 100

Arrangement for filtering compressed pulse-code-modulated signals

L. D. J. Eggermont E

Digital filter for filtering nonuniformly quantized pulse code-modulated signals formed by a sequence of code words each comprising a segment number $s(i)$ and a mantissa number $m(i)$. This digital filter comprises a modifying device to which the mantissa numbers $m(i)$ are applied for generating modified mantissa numbers $E(i)$. The numbers $E(i)$ are each multiplied by the magnitude of a filter coefficient, for generating first product numbers $z_1(i)$. Thereafter these numbers $z_1(i)$ are each multiplied by a number $2^{B(i)}$, $B(i)$ being either equal to $s(i)$ or equal to $s(i) - 1$, so that second product numbers $z_2(i)$ are obtained which are applied to an accumulator.

---

4 231 101

Digital filter arrangement for non-uniformly quantized PCM

L. D. J. Eggermont E

Digital filter for filtering non-uniformly quantized pulse code modulated signals formed by a sequence of code groups $x(i)$, each comprising a polarity bit, a segment number and a mantissa number. This digital filter arrangement comprises a generator which cyclically generates a series of sequentially occurring auxiliary numbers each comprising a polarity bit and an address code; a first storage medium for storing the code groups $x(i)$ and controlled by an address computation circuit to which the auxiliary numbers are applied; a second addressable storage medium to which a segment number $s(i - h)$ a mantissa number $m(i - l)$, supplied by the first storage medium, as well as a polarity bit supplied by an exclusive OR-gate are applied as the address code. This polarity bit is derived by the gate from the polarity bit $p(i - l)$ of the code groups $x(i)$ and the auxiliary numbers polarity bit. This second storage medium

contains positive numbers z_p and negative numbers z_n which represent the expanded values of all possible code groups $x(i)$. A positive number z_p is read from the storage medium if the polarity bit is positive and a negative number if the polarity bit is negative. The numbers read from the storage medium are applied to an accumulator which is coupled to an address code output of the generator.

---

4 233 261

Method and device for manufacturing information carriers

A. Mijnheer E

Method of and a device for manufacturing information carriers from a thermoplastic material, where a thin information carrying die is heated by bringing it into contact with a relatively thick, heated surface and is subsequently separated from the heated surface and pneumatically pressed against the thermoplastic material by a layer of air that thermally isolates the die from the heated surface so that the die may rapidly cool.

---

4 233 385

Method and apparatus for liquid electrostatic development of charge images on a tape-like record carrier

H.-D. Hinz H
U. Rothgordt
F. Schinke

A method of liquid development of charge images formed on a surface of a tape-like record carrier, for example by an electrostatic printer. The record carrier is simultaneously sprayed with developer liquid in two flows which are directed towards each other. As a result, two separate, uniform and oppositely directed flow zones meeting at one common turbulent flow zone are obtained. Both during pre-development and final developement the charge images are brought into contact with a large quantity of fresh developer liquid.

---

4 233 489

Method of and device for plasma-MIG welding

W. G. Essers E

Plasma-MIG welding wherein a plasma arc is initially established between a non-consumable electrode and a workpiece to initiate and sustain a non-deviated plasma flow therebetween, welding wire is supplied along a path generally parallel to but laterally displaced from the plasma flow path, and a MIG arc is thereafter struck between the welding wire and the workpiece, whereby the plasma flow is laterally displaced by and drawn to the MIG arc to substantially concentrically surround the same.

---

4 233 502

Opto-electronic focusing error detection arrangement

G. Bouwhuis E
T. J. Hazendonk

An opto-electronic focusing-error detection arrangement is described for the detection of a deviation between a radiation reflecting surface and the plane of focusing of an objective system. In the path of the beam which is reflected by the surface a beam-splitting element is disposed and behind said element two radiation-sensitive detectors are arranged which are each associated with one of the subbeams formed by the beam-splitting element. The detectors are grating-shaped detectors and are effectively divided into two detector sections by selection devices, the bounding line being adjustable. A focusing error signal is obtained which is highly independent of a positional error of the radiation-sensitive detection system relative to the beam axis.

4 233 617

Field effect transistor with insulated gate electrode

F. M. Klaassen E
J. A. Appels

A field effect transistor of the V-MOST type in which the channel region comprises a more highly doped part which adjoins the source zone and a lower doped part which surrounds said region, said channel region adjoining the surface and surrounded by an insulation diffusion. The lower-doped part is depleted from the *PN*-junction with the low-doped drain region up to the surface at a voltage which is lower than the breakdown voltage.

---

4 233 684

Arrangement for decoding a signal encoded by means of adaptive delta modulation

L. D. J. Eggermont E

Arrangement for decoding a compressed delta modulation signal, wherein, to simplify the analogue output filter, an interpolating digital filter having a signal-independent pulse response is included in the dynamic expansion circuit.

---

4 234 763

Feeding bridge with d.c.-compensation for both directions of the feed current

E. C. Dijkmans E
B. IJff
A. H. J. Reuvekamp

A bridge circuit comprising an isolation transformer having six primary windings which are connected so that the voice currents do and the supply currents do not generate, for either direction of current flow, a resulting flux in the core of the transformer. The bridge comprises two transistor circuits which prevent the voice currents from flowing through a supply source connected to the bridge and also balance the bridge circuit so that the influence of longitudinal noise signals which are produced in a transmission line connected to the bridge circuit are suppressed. The direction of the feed current in the transmission line can be reversed by means of the transistor circuits.

---

4 234 778

Method of and welding torch for arc welding

G. A. M. Willems E
G. W. Tichelaar

Plasma-MIG welding in which a thermally ionizable gas stream is flowed through a nozzle non-consumable electrode having a central orifice and a surrounding annular opening toward a workpiece and is thereby split into a central gas column enveloped by an annular gas shield. A consumable electrode is fed through the central gas column toward the workpiece, with the establishment of a MIG-arc therebetween. A plasma arc is then spontaneously established by means of the MIG-arc between the nozzle non-consumable electrode and the workpiece. The central plasma gas column is accelerated by constriction of the annular gas shield down-stream of the nozzle non-consumable electrode.

---

4 234 807

Ladder device with weighting factor adjusting means

L. J. M. Esser E
L. G. M. Heldens

A ladder device comprising at least three ladder sections connected in cascade and weighting-factor adjusting means coupled to said ladder sections and which comprises first and second field effect transistors (FET) having channel regions with different predetermined length/width ratios that define a weighting factor coefficient. A further fine adjustment of the weighting factor is achieved by adjusting the gate voltages of the FETs.

4 236 045

Electric lamp

R. J. Q. van den Plas E
P. Hokkeling

In electric lamps a type of glass is used for the envelope which, in many cases, has a coefficient of expansion which differs considerably from that of the current-supply conductors. Therefore, special measures have to be taken to seal the lamp envelope in a vacuum-tight manner around the current-supply conductors. The invention provides a simple, vacuum-tight seal of a lamp envelope, which seal consists of a metal plug which is sealed both to the glass of the lamp envelope and to the current-supply wires. The metal plug of 100 parts by weight of a first metal (tin and/or lead) and 0.05–1 parts by weight of a second metal (titanium, zirconium, hafnium, niobium, tantalum, and vanadium) has a strong adhering power and a large ductility.

---

4 236 173

Apparatus for reading a disc-shaped record carrier having a PAL color television signal

M. P. M. Bierhoff E
A. H. Hoogendijk

An apparatus for reading a disc-shaped record carrier on which per track circumference one picture of a color television signal is recorded. The signal being read is converted to a standard PAL color television signal. The apparatus comprising a command device which by applying command signals to control means can produce a jumpwise radial change in the scanning position on the record carrier, so as to enable the scanning sequence of the recorded pictures to be changed. In order to maintain a standard PAL color television signal in the case of a changed scanning sequence, the apparatus comprises a correction device for the chrominance signal. This correction device is controlled by the command device and comprises a mixing circuit for mixing the chrominance signal with a reference signal of twice the chrominance subcarrier frequency during time intervals prescribed by the command device, and phase correction means for introducing a phase shift of 0°, 90°, 180°, or 270° in the chrominance signal in accordance with a pattern defined by the command device.

---

4 236 175

Converter circuit and monochrome picture display device comprising such a converter circuit

H. H. H. Groothuis E

To display bivalent color signals in a monochrome picture display device so that the original color remains recognizable by choosing a suitable luminance ratio, and colors having a low brightness are displayed with a sufficient degree of visibility, the combination of a sum signal and a correction signal is formed from the bivalent color signal, wherein the color differences are derived from the sum signal and the correction signal ensures the basic luminance.

---

4 236 225

Data buffer memory of the first-in, first-out type, having a variable input and a fixed output

P. G. Jansen E
J. L. W. Kessels
B. L. A. Waumans

A data buffer memory of the first-in, first-out type, having an input bus by which data are supplied to the buffer, and a fixed output from which data are transferred from the buffer. Each section of the buffer includes logic whereby a variable input location can be selected. Status signals are used to determine, in cooperation with signals supplied from outside the buffer, the location where data are to be written in the buffer and when these data in the buffer must be shifted to the output.

4 236 226

Magnetic domain device

E. H. L. J. Dekker E
U. E. Enz
J. Haisma
K. L. L. van Mierloo

A magnetic device comprising at least one thin domain layer of a magnetizable material which has an easy axis of magnetization which is substantially normal to the surface of the layer and in which magnetic domains are propagated under the influence of a bipolar current, for example an alternating current, by a pattern of electrically conductive material with which the layer is provided. Elements are furthermore present which cause an asymmetry force and thus determine the direction in which the domains are propagated. The electrically conductive material and the elements are present in a single pattern which is constructed from at least a layer of magnetic material.

---

4 236 232

Optical read apparatus for reading a disc-shaped record carrier

G. L. M. Jansen E
L. G. H. Bovee

An optical read apparatus for reading a disc-shaped record carrier. A read element is mounted on an adjusting member by means of which the read element is radially movable. Moreover, there is provided a deflection element for correcting the radial position of the scanning spot. In order to enable an arbitrary information unit on the record carrier to be read with a short searching time, the drive of both the adjusting member and the deflection element is controlled. In order to enable the adjusting member to be moved in a rapid and controlled manner there is provided a position indicator, which indicates the instantaneous radial position of the read element. Moreover, there is included a correction circuit which eliminates the adverse effect of inaccuracies in the position indicator on the searching time.

---

4 237 400

Low-pressure discharge lamp with tortuous discharge path

G. A. Wesselink E
H. Roelofs
C. H. M. van Bommel

Low-pressure discharge lamp having a discharge space, limited by an elongate lamp vessel, electrodes at one end of the vessel in the discharge space between which electrodes a discharge takes place during operation of the lamp. The lamp vessel has partitions to divide the discharge space into chambers each extending substantially in the length of the discharge vessel which chambers communicate with one another via at least one opening. The chambers are sequentially passed through by the discharge, the opening between at least two chambers being at least partly wedge-shaped, the minimum width being located nearest to an electrode whose discharge path extends itself after ignition.

---

4 237 480

Television camera with pick-up tube mounting means

A. J. J. Franken E
W. W. J. Degger

A television camera in which a color splitting prism and a number of pick-up tubes are positioned, with respect to each other in a holder which comprises an annular seat for optical alignment of each pick-up tube, the relevant pick-up tube being pressed onto said seat, by way of an adapter ring connected thereto transversely of the tube axis, by means of a resilient pressure member which acts in the direction of the tube axis. The diameter of the parts of the seat and the adapter ring decreases in the pressing direction. The pick-up tubes fit in the holder in a locating manner and, after being pressed, do not exhibit play with respect to each other, so that variations of the position with respect to each other in reaction to shocks, and inherent picture registration errors, are avoided.

PHILIPS TECHNICAL REVIEW

VOLUME 39, 1980, No. 9

COLATH, a numerically controlled lathe for very high precision

T. G. Gijsbers

The ability to make 'optically smooth' surfaces by machining metals, alloys and plastics has greatly broadened the horizons of applied optics. This new capability is largely a result of the development of numerically controlled precision lathes, which bring greater versatility to production methods and also cut costs. COLATH, the numerically controlled lathe described below, has been developed through the efforts of a team of about ten people at Philips Research Laboratories. A particularly powerful and trend-setting stimulus for the work carried out on the machine is provided by the Philips optical sector. A typical example is to be found in an article by J. Haisma and his colleagues (referred to here in a footnote) on the manufacture of 'bi-aspherics', i.e. lenses with two aspherical surfaces, often strongly curved.

The development of precision machining at Philips

Our company's long-standing interest in the machining of all sorts of articles and components to give their surfaces a dimensional accuracy and smoothness that will satisfy the highest — i.e. optical — requirements dates back to a period just after the Second World War when Philips Research Laboratories started this work [1]. To obtain such precision it is necessary to devote a great deal of attention to three important aspects: firstly the control of all the movements that the workpiece and cutting tool may make with respect to one another — characteristics of the machine, secondly the cutting tool itself, and thirdly the choice of material to be machined (metals, alloys, plastics); in this article we shall only deal with the materials in passing. Even before 1960 considerable progress had been made both with the control of machine movements and with the cutting tool. By that time L. M. Leblans had considerably improved the operation of lathes and milling machines by his development of the 'vacuum head' [2]. In this design of main spindle for machine tools he had two reasons for employing a vacuum: firstly so that the lubricant could be continuously transported through the bearing clearances (with a height of only a few microns), to ensure continuous hydrodynamic lubrication and, secondly, to apply to the main spindle a constant axial force to press it against a high-precision thrust bearing. This considerably increased the machining accuracy in the axial direction. (At the time the tolerance that could be achieved in the radial direction was about 1 μm.) Accuracy in the axial direction is of the utmost importance in the preparation of surfaces for lenses, for example. At roughly the same time, in about 1960, precision turning and milling were also being used in the manufacture of parabolic reflectors, various metallic reflectors, resonator cavities, and similar devices.

At much the same time Leblans was improving the manufacture of the cutting tool, which had a single-crystal diamond tip. He designed and built a special grinding machine that could be used for grinding natural diamond accurately to the desired shape. A method was also developed for brazing the diamond tip to the tool.

Ir T. G. Gijsbers is with Philips Research Laboratories, Eindhoven.

[1] H. Rinia and P. M. van Alphen, The manufacture of correction plates for Schmidt optical systems, Philips tech. Rev. 9, 349-356, 1947/48.

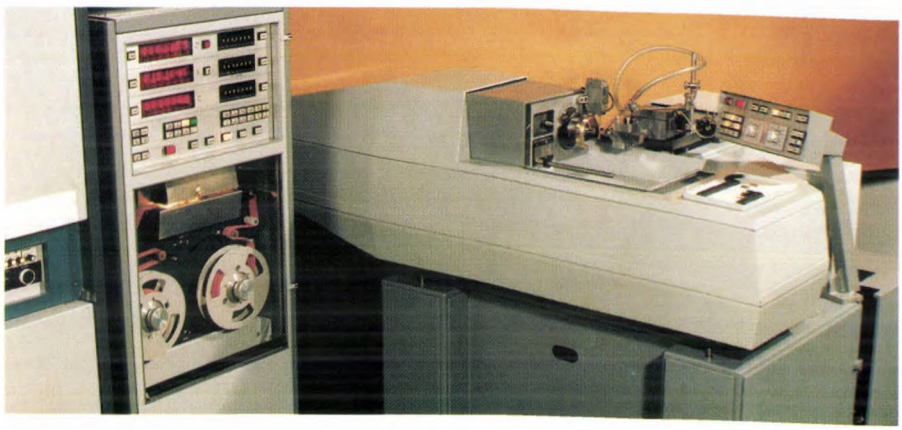
[2] L. M. Leblans, A high-precision lathe headstock, Philips tech. Rev. 19, 68-69, 1957/58.

The result of all this work was that the special department for surface machining at Philips Research Laboratories had to expand considerably. One of its most important early activities was the application of ultra-fine turning with a diamond tip for facing up memory discs. The flatness of the surface determines the minimum distance at which a read/write head can be mounted above a memory disc. Signal transmission improves as the head is brought closer to the disc, but there is then a higher probability that the rapidly rotating disc and the stationary head might touch and cause damage. The flatness obtained was so good that a compromise distance of about 2 μm could be used.

The surface-machining department also assembled a number of new machines, some of them highly specialized, from parts of existing lathes and vacuum heads of various designs. A good example of this is the much-used spherical-turning machine that can produce optically smooth spherical surfaces (concave and convex, maximum diameter about 10 cm).

In the sixties our machines were considerably improved as a result of the concentrated study on hydrostatic bearings and hydraulic servomotors carried out at that time by the mechanical research group at Philips Research Laboratories. This research resulted in the design and construction of various new headstock and carriage arrangements that include very stiff hydraulic bearings and extremely accurate hydraulic drive systems.

The culmination of all this work came in 1967 and 1968 when H. J. J. Kraakman and J. G. C. de Gast designed and built a precision lathe [3]. In this new lathe, which was constructed entirely by Philips, all the bearings are hydrostatic. There is therefore no possibility of metal-to-metal contact between the moving parts of the lathe. The carriages are driven by linear hydraulic motors (or 'actuators').



a

b

Fig. 1. *a*) A general view of COLATH, the numerically controlled precision lathe developed at Philips Research Laboratories. *Left:* a computer control unit (CNC) with a control panel at the top containing the numerical display units for the positions of the toolholder and headstock carriages. There is a second control panel on the lathe itself (*right*); one use of this panel is for accurately positioning the tool at the start of a machining program. *b*) These two bi-aspheric lenses were made on COLATH. They can be used as an objective in a 'VLP' player.

A linear hydraulic motor drive has two great advantages: the drive is direct and exceptionally 'stiff', much stiffer in fact than the conventional mechanical drive; the second advantage is that the drive mechanism has no periodic structures (such as screw threads). This greatly reduces the possibility of 'ghosts' appearing in the machined surface.

The main spindle is driven by a vane-type motor, a rotary hydraulic motor specially designed and constructed by Philips. The lathe can only be controlled manually; the speeds at which the carriages can travel (in the direction of the axis of rotation and perpendicular to it) are adjusted by means of variable flow-control valves. The form accuracy that can be obtained is 1 μm and the surface finish to be achieved can be described as 'optically smooth'. Kraakman and De Gast's lathe has proved to be extremely suitable for geometrically simple work such as turning the surfaces of cylinders, bushes or planes to a mirror finish. The manual control of the carriages does have its limitations, however. Workpieces of a more com-

plicated shape, like the aspheric lenses now in such demand, cannot be machined on this lathe.

In the early seventies we began to think of computer control for a precision lathe of this kind instead of manual control with its associated limitations. A precision lathe of this type with Computer Numerical Control (CNC) was completed and put into service during 1978. The new machine (*fig. 1*), which we called COLATH, allows us to make workpieces of any *arbitrary* rotationally symmetric shape. The form accuracy attainable is better than 0.5 μm for a maximum workpiece diameter of 20 cm. Under favourable conditions the surface roughness is 0.02 μm, which is undoubtedly 'optically smooth'.

The demand for this new precision lathe came mainly from those working in 'optomechanics', a relatively new field in which precision-machining operations are used to provide the exceptionally high quality of finish necessary today for optical components.

Our first experiences with COLATH in such activities were gained in making bi-aspheric plastic lenses. A workpiece of this kind (fig. 1*b*) usually has to satisfy the exacting requirement that the natural diffraction of the light at the edge of the lens should be the factor that limits the resolution. Any deviations from the ideal shape of the refracting surfaces of the lens should therefore have less effect. Optical components that meet this requirement are said to be 'diffraction-limited'. The deviations from the ideal lens shape are then so small that the optical path differences do not exceed $\frac{1}{4}\lambda$ (Lord Rayleigh, 1878).

As well as its shape, the surface topography of the lens is also very important. The roughness of the refracting surfaces should be low enough for them to be considered as 'optically smooth'. The measure of this quality is not a concept that can be defined very precisely; a practical rule often used is 'the maximum intensity of the central diffraction image should be at least 80% of its theoretical maximum intensity'. The surface is then said to be 'optically smooth', i.e. 'perfectly' transparent (or reflecting). In our case this quality is difficult to obtain because the lathe tool generates a regular pattern on the surface. The shape of this pattern is determined by the radius of the diamond and the lateral movement of the tool or workpiece in each revolution. *Fig. 2* is a diagram of such a pattern (see also the illustration on page 183 in No. 6/7 of the current volume of this journal). The pattern does not seem to cause any undesirable diffraction-grating effects of practical significance, provided the profiling (*P* in fig. 2) does not go any deeper than 0.02 μm. The surface is then 'perfectly' transparent or reflecting to visible light.

Variations in the shape of the lens can cause two different errors in the wavefronts: errors that vary slowly across the surface of the lens (errors of form) and errors that vary rapidly across it (roughness errors). If there are errors of form, light disappears from the central diffraction image and appears in the first diffraction rings around it. For visible light the 'diffraction-limited' requirement is equivalent to a form error less than about 0.1 μm over the entire light-refracting surface. In roughness errors the light is scattered relatively uniformly in a broad cone.

Because of their periodicity, the residual machining marks caused by the tool can produce a fairly clearly defined light ring. The pitch *S* in fig. 2, the distance travelled by the cutting tool in one revolution, is of direct optical significance. If this distance is approximately equal to the wavelength of the light used, the surface will scatter this light appreciably, and this means that the energy efficiency of the optical component is low (i.e. there will be little energy radiated in the direction of the zero order). From a technological point of view the most convenient value of *S* at present is somewhere between 5 and 10 μm. This is satisfactory for visible light, but not so good for the infrared part of the spectrum. Since there has been much interest in infrared in the last few years, it will be necessary to investigate the machining process more closely to try to ensure that the most convenient value of *S* is as different as possible from the wavelengths to be used.

As mentioned earlier, the tool has a tip of a sharp and well-defined single-crystal diamond. This limits the materials that can be machined to the non-ferrous

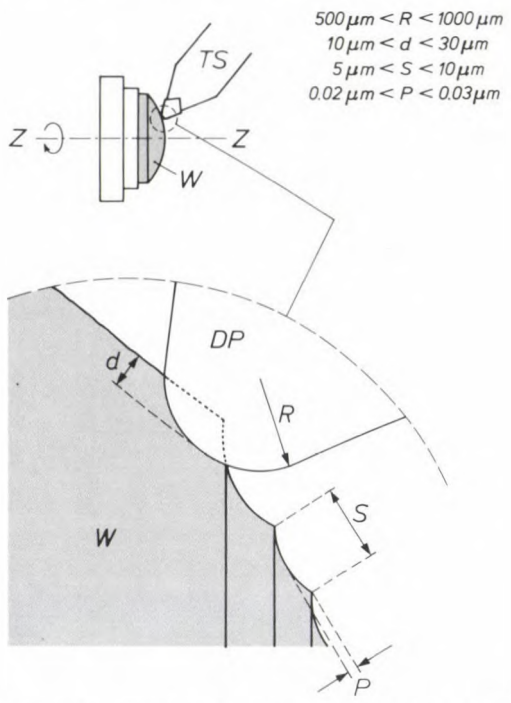


Fig. 2. Turning a workpiece *W*, which rotates about the axis *ZZ*. A groove pattern (see enlargement) with pitch *S* and profile *P* is formed on the surface. The diamond tip *DP* (radius of curvature *R*) is brazed to the shank *TS*. The thickness of the chip that is removed is determined by the depth of cut *d*. The dotted line indicates the cross-section of the chip. Typical values of *R*, *d*, *S* and *P* for machining on COLATH are shown in the figure.

[3] H. J. J. Kraakman and J. G. C. de Gast, A precision lathe with hydrostatic bearings and drive, Philips tech. Rev. **30**, 117-133, 1969.

metals and their alloys (aluminium, copper, silver, brass, etc.) and the machinable plastics. Iron in the material would dissolve traces of carbon from the diamond and this would soon blunt the cutting edge. Furthermore, when a material containing iron is being machined, the diamond is liable to crumble, owing to the presence of hard inclusions (e.g. carbides) in the material. A hard material of this kind, e.g. for a mould for glass pressing, has to be ground and this might also be done on COLATH.

The first plastic lenses made with COLATH were examined on an optical bench and were found to be of the desired 'diffraction-limited' quality.

It will be obvious that COLATH is very useful not only for optomechanical work but also has much to offer for other precision-engineering applications; one example is the manufacture of components for air bearings.

The next section gives a discussion of the general design of COLATH and how the machine is used. In the sections following this the hydraulic system, the laser measurement system, the numerical control and the control loops of the lathe will be discussed in greater detail. The final section gives details of workpieces that have already been made on COLATH. These are mainly aspherical optical components — for possible application in the 'VLP' system [4] and for X-ray equipment and infrared instrumentation. General technical information about the lathe and its performance have been summarized in tables.

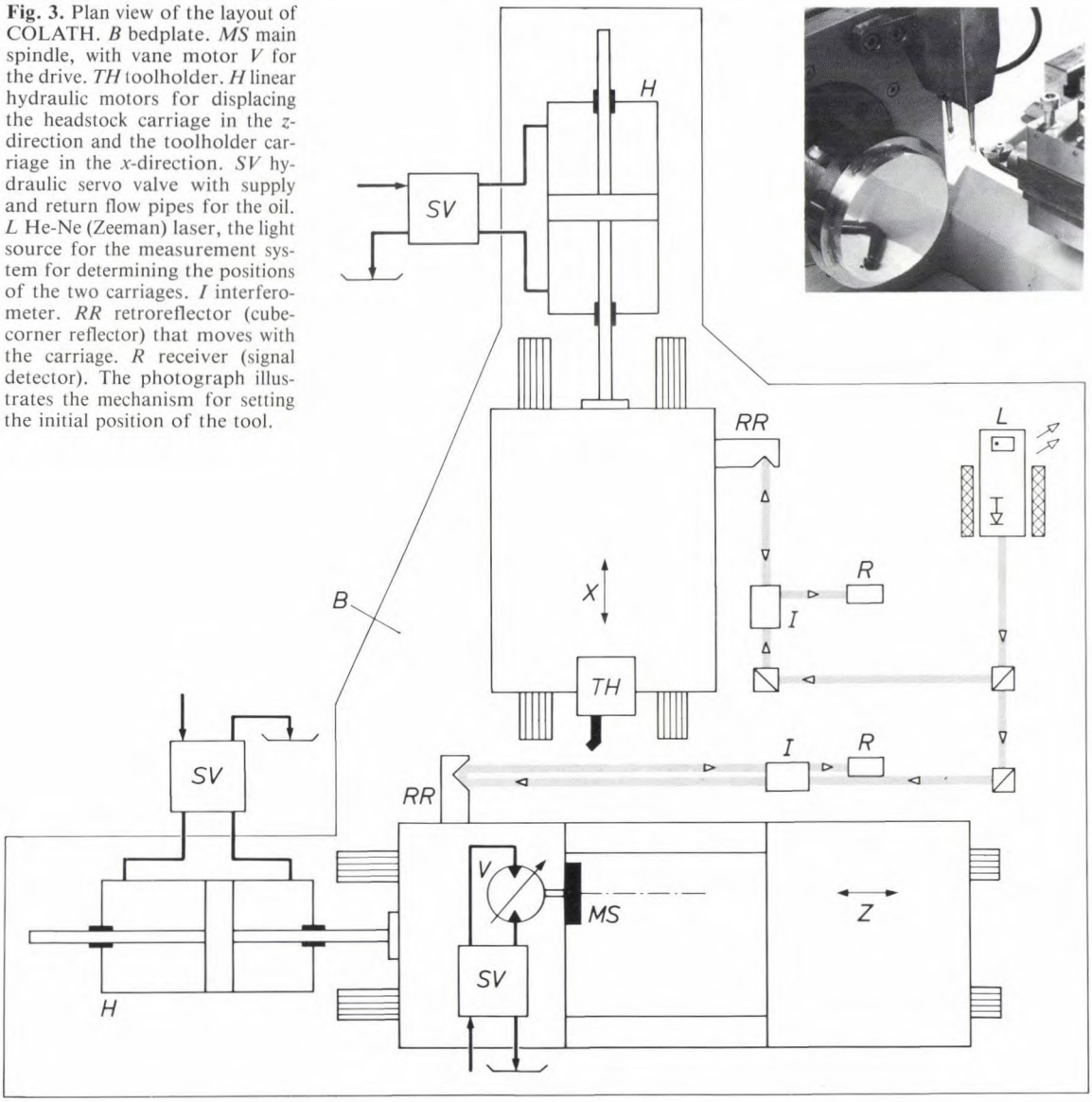


Fig. 3. Plan view of the layout of COLATH. *B* bedplate. *MS* main spindle, with vane motor *V* for the drive. *TH* toolholder. *H* linear hydraulic motors for displacing the headstock carriage in the *z*-direction and the toolholder carriage in the *x*-direction. *SV* hydraulic servo valve with supply and return flow pipes for the oil. *L* He-Ne (Zeeman) laser, the light source for the measurement system for determining the positions of the two carriages. *I* interferometer. *RR* retroreflector (cube-corner reflector) that moves with the carriage. *R* receiver (signal detector). The photograph illustrates the mechanism for setting the initial position of the tool.

General features of the machine and its use

Design

A plan view of COLATH is shown in *fig. 3*. Principal data and performances of the lathe are listed in *Table I*. As in Kraakman and De Gast's machine [3], which will be designated simply by its type number (PHM1000) from now on, the toolholder is located on a carriage that can move in the x-direction and the headstock is mounted on a carriage that can move in the z-direction. These two structures are mounted *independently* of one another on the heavy bedplate of the lathe, and because of this have hardly any effect on one another. The headstock carriage is much longer than in the PHM1000. This has the advantage that it is far less easily tilted, since its bearings are further apart.

Like the PHM1000, COLATH has hydraulic motors: two linear actuators for moving the two carriages and a rotary vane motor to drive the main spindle. The vane motor is built in as an integral part of the structure of the headstock unit. The motors are a little different from those used in the PHM1000. The differences, which will be examined in greater detail in the next section, have resulted in greater stiffness of the hydraulic system and consequently a faster response from the control loops.

The movements of the carriages in the x- and z-directions are continuously controlled, in the same way but independently of one another, by a CNC unit. The principle of the control system is illustrated in *fig. 4* and is the same for both directions. This highly sensitive system gives the new lathe essential advantages over the old one. It has two laser interferometers that continuously measure the positions of the carriages (resolution about 16 nm) and feed back the measured data to the comparator unit of the appropriate control circuit.

Table I. Leading technical data and performance of COLATH.

Clamping diameter	max. 220 mm
Turning dimensions	max. 200 mm diameter (x) max. 200 mm length (z)
Carriage velocity ($\dot{x}$ and $\dot{z}$)	1.8-180 mm/min
Carriage acceleration ($\ddot{x}$ and $\ddot{z}$)	1/32-1/16 μm/ms^2
Speed of main spindle	0-1000 rev/min
Contour	free choice
Form error [a]	<0.5 μm
Surface roughness	about 20 nm (peak-to-valley)

[a] In theory $(\Delta x^2 + \Delta z^2)^{\frac{1}{2}}$, the distance between a measured point on the actual contour and the corresponding point on the theoretical contour.

Both the carriages and the motors have hydrostatic bearings throughout, so that there is no possibility of any kind of dry friction, the friction that causes the highly undesirable 'stick-slip' in movement. The self-adjusting couplings between the carriages and the piston rods of the linear actuators also have hydrostatic bearings (this was not so in the PHM1000). A well-known advantage of hydrostatic bearings compared with air bearings is that their stiffness is higher; this clearly is better for machining accuracy [3].

Preparation for turning

Before the automatic execution of a workpiece program can be started, the tool tip must be set to a position accurately known in relation to the axis of the machine. This is done by means of a 'mechanical zero' on the machine (fig. 3). The uncertainty associated with this zero is less than 0.5 μm.

The numerical information for turning a workpiece is processed by means of special computer programs in a separate minicomputer and then transferred to a tape. The punched tape can be fed to the CNC unit of

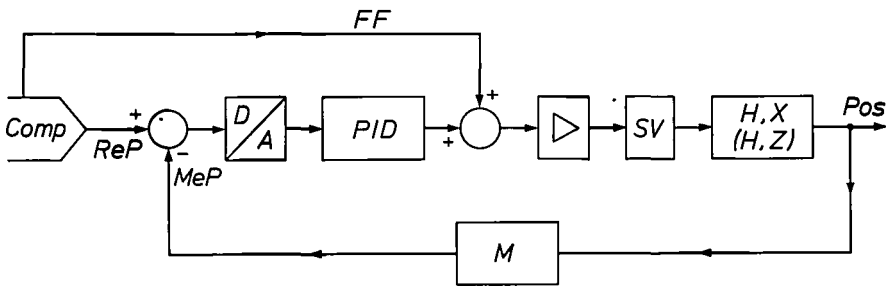
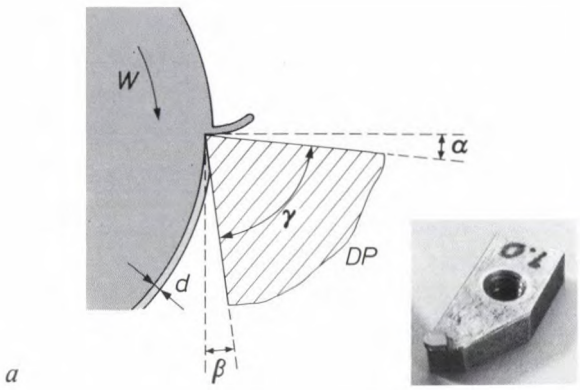


Fig. 4. Block diagram for the control of the displacement of a carriage. *Comp* computer. *ReP* required position. *MeP* measured position. *Pos* controlled position. The control loop illustrated can be used both for the carriage with the headstock and for the carriage with the toolholder. *D/A* digital-to-analog converter. *PID* electronic controller with proportional, integrating and differentiating action. *SV* hydraulic control valve. *H,X(H,Z)* linear actuator with the carriage for *X* movement, or the carriage for *Z* movement, connected to it. *M* laser interferometer. *FF* parallel-input signal, proportional to the velocity of the required displacement.

[4] K. Compaan and P. Kramer, The Philips 'VLP' system, Philips tech. Rev. 33, 178-180, 1973.

the machine. When it is preparing the tape the minicomputer takes account of a 'tolerance zone', extra input data that sets the tolerance limits within which the computer program should approximate to the theoretically desired shapes of the workpiece.



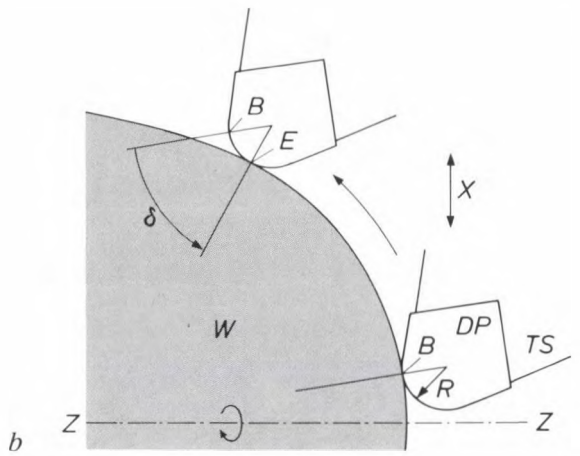


Fig. 5. *a*) A diamond-tipped tool for COLATH. The drawing shows, in vertical cross-section, how the diamond tip removes a chip with a cutting depth d. α rake angle, usually close to zero. β clearance angle, usually about 5°. γ wedge angle, 85°. *b*) Illustration of the 'unrolling' of the cutting edge of the diamond during the turning of a curved contour. δ 'angle of roll' relating to the turning of the section between B and E. *DP* diamond tip. R radius of the cutting edge (total useful arc approximately 120 degrees). *TS* tool shank. W workpiece. ZZ axis of rotation.

The cutting tool

The cutting tools have a tip of single-crystal natural diamond; they have been made at Philips Research Laboratories for more than 20 years. The first operation in the process is to braze a chip of diamond on to a molybdenum shank. The diamond chip is then ground to the desired circular shape in the special grinding machine mentioned earlier. The radius of curvature of the diamond tips made in this way has a value between 0.5 and 1 mm. The grinding is carried out very accurately, so that the tolerance in the radius of curvature is less than 0.2 μm over the entire 'angle of roll', which can be as much as 60 degrees. The name 'angle of roll' is used because part of the cutting edge of the diamond tip is 'unrolled' when curved surfaces are turned (*fig. 5*). Because of this it is very important for the form accuracy of the desired surface that the radius of curvature of the tip is constant. The cutting quality of the tool must also be high and must also remain constant over this angle of roll. The photograph in fig. 5*a* shows a ground diamond tip on the molybdenum shank. The tool is checked after it is made to ensure that the tip has the correct roundness.

For most materials that are machined a rake angle of zero and a clearance angle of 5 degrees are satisfactory. The diagram in fig. 5*a* shows how these two angles are defined. When the diamond tip is being ground, the flank of the tool is given a conical shape so that the clearance angle remains constant for any point on the cutting edge.

Working with a numerically controlled precision lathe

The technician using a lathe such as COLATH soon realizes that even though there is an automatic control system his expertise is essential (*fig. 6*). Although most of the manual operations are no longer necessary during the actual cutting process, it is still vitally important to have continuous expert and meticulous human checking and supervision, particularly in the preparatory operations for machining. In the clamping operations, for example, any deformations must be less than 0.1 μm. It goes without saying that special attention must be paid to the metrology, which is 10 to 100 times more accurate than the conventional methods based on micrometers and vernier calipers. This high measurement accuracy is wasted if other aspects of the work, such as the clamping operations, are not carried out with the utmost care.

For clamping up, the user of the lathe will often have to devise *and* make his own special aids for the job. His skill must be sufficient to enable him to take the proper corrective action if he notices something out of the ordinary, for example a coloration of the surface being machined. This is by no means as simple as it seems, for the uninitiated observer often sees no colour at all. The technician then knows that the tool is just a little too sharp. However, if he has had a great deal of experience, the very details of such a situation as this would tell him *not* to intervene. Just continuing the turning will bring the tool to the correct sharpness. Such decisions are only reached as the result of careful and concentrated 'cooperation' between the technician and the automatic system. A mark of the expert technician is that he knows exactly when the top face of the tool should be 'smoothed' again.

To sum up, machining still requires the greatest care and attention, in spite of the automation, primarily because the machining accuracy demanded is so extremely high. Because of the high precision the technician also has to make due allowance for the temperature distribution in COLATH; at the least he should know when the temperature distribution has become stable enough for the machining operations to be performed with sufficient reproducibility.

In ordinary lathes vibration may occur during the work, with the result that the quality of the finished product is unsatisfactory. The COLATH user should be even more attentive to this; a store of personal know-how on vibration damping is a great help here.

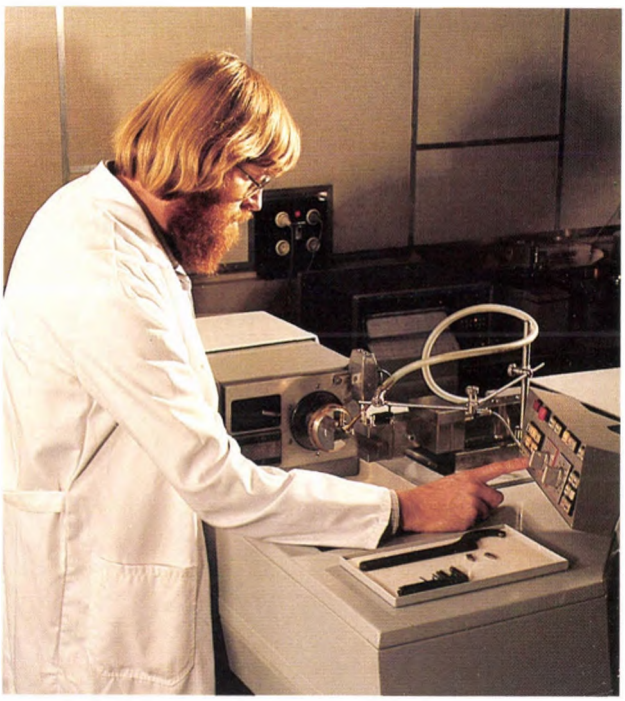


Fig. 6. The photograph shows a technician working with COLATH. The small manual control panel on the machine (*right*) is useful for work that has to be kept under direct observation.

The handling, cleaning and packaging of the workpieces are all delicate operations that require a great deal of skill and care if the high quality is not to be lost.

The hydraulic system also requires the proper attention; it has its own requirements for running in and maintenance (the oil must be properly de-aerated and the filters must be cleaned at the correct intervals, for example). This system does not function independently of other major parts of the equipment, such as the control loops.

The COLATH technician obviously needs to have a good general picture of these functional interrelationships inside the lathe. He also needs to have a basic understanding of the programming methods, since the good technician will also want to be able to take action here if necessary, either on his own initiative (e.g. to correct the punched tape) or in association with the part programmers.

If COLATH is to be operated correctly and we are to take full advantage of all its special features, then we need people with a broad technical understanding and a great deal of detailed know-how. They will also need the right sort of disposition and the ability to do the job 'in cooperation with' an extremely sensitive and highly automated system — a relatively rare combination of man and machine.

The hydraulic system

The hydraulic system of COLATH is illustrated in *fig. 7*. It consists of a pump unit (*P*) with a reservoir, two linear units (*X* and *Z*) for moving the carriages, and a rotary unit (*MS*) for driving the main spindle, as well as all the hydrostatic bearings (*HB*) in the various units.

The oil supply

The unit *P*, which contains four independently operating pumps and a cooler, supplies oil to the two linear actuators (H_1 and H_2, pump P_1) and the vane motor (*V*, pump P_2), and also to all the hydrostatic bearings (*HB*, pump P_3). The vane motor has a *separate* pump — this was not so in the PHM1000 — to prevent interaction between the motors H_1, H_2 and *V*. The pump P_4 of *P* pumps the oil through the cooler and back to the reservoir; more will be said about this later. To keep the temperature of COLATH sufficiently constant during operation, considerable cooling capacity is available — this too will be discussed in more detail later.

In the oil-supply lines between the pumps and the motors there is an electromagnetically operated control valve (*SV*). This valve, which is also known as a 'servo valve', controls the magnitude of the volume flow of oil and also, between valve and motor, its direction.

The volume flow of the oil is the quantity directly proportional to the linear speed (or the speed of rotation if appropriate) of a motor (in quasi-static applications). In principle, the control valves *SV* determine all the forward and backward movements of the motors. The speeds are continuously variable about a central position of the valves; the central position corresponds to a standstill. This type of control valve, known as a 'throttling directional-control valve', can be considered as a unit that converts an analog electrical signal into an analog hydraulic signal.

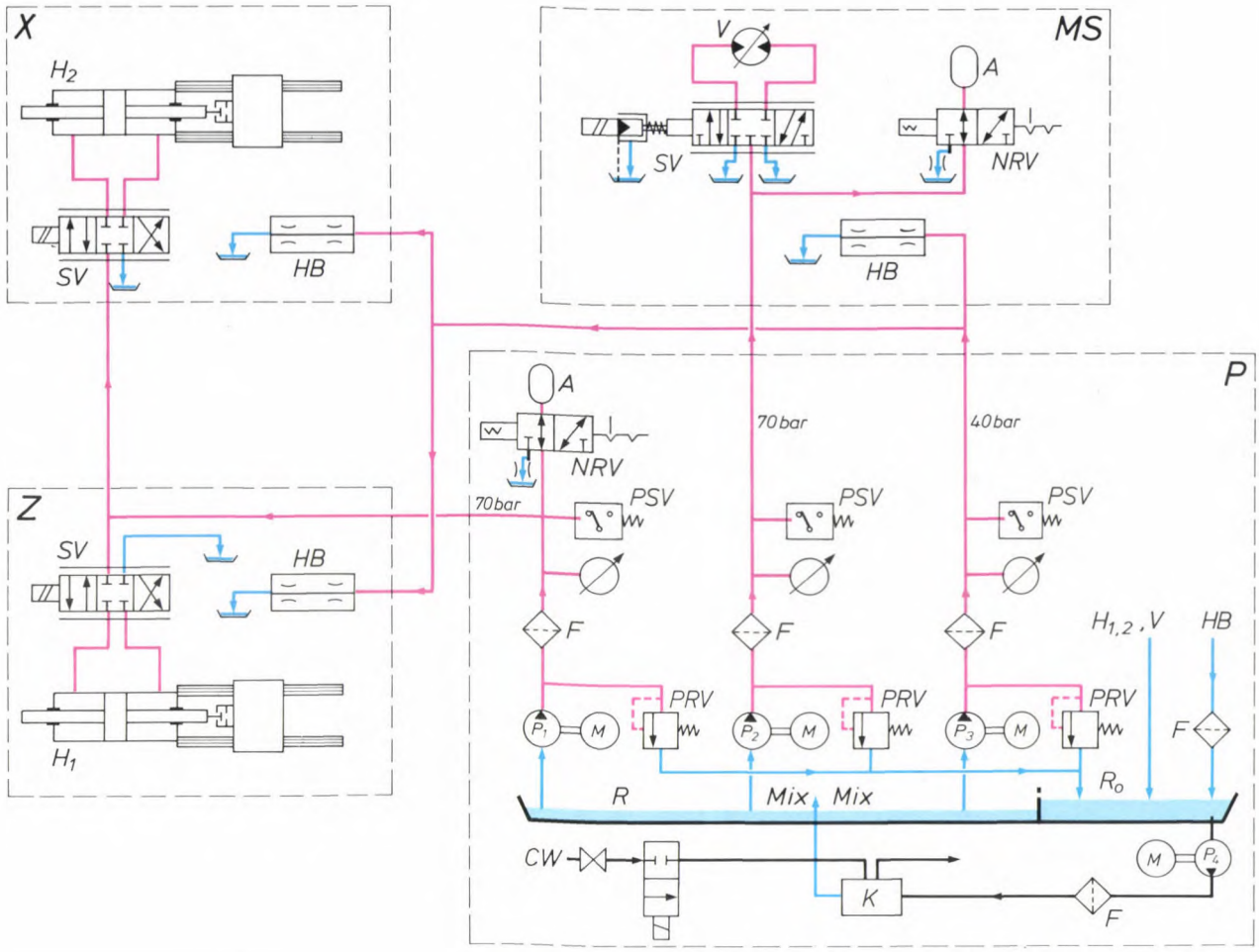


Fig. 7. Diagram of the hydraulic system of COLATH. *P* pump unit. *X* linear drive unit for moving the toolholder carriage. *Z* linear drive unit for moving the headstock carriage. *MS* rotary unit for driving the main spindle. *HB* hydrostatic bearing. *SV* electromagnetic control valve. $H_{1,2}$ linear actuator. *V* vane motor. $P_{1,\ldots,4}$ oil pump. *M* electric motor. *PRV* pressure-relief valve. *A* accumulator. *NRV* non-return valve. *PSV* hydraulic switch. *F* filter. R_o overflow reservoir. *K* water-cooling system. *R* reservoir. *CW* cooling water. More information about the hydraulic symbols used in this diagram is given in ISO standard 1219-1976 'Fluid power systems and components — Graphic symbols' published by the International Organization for Standardization (ISO) and also in the latest edition of the Hydraulic Handbook (published by Trade and Technical Press, Morden, Surrey, England).

The oil-supply lines contain pressure-relief valves (*PRV*) to ensure that the operating pressure does not exceed the desired values; they also contain filters (*F*) to stop the oil from being contaminated. If there is a dangerous pressure drop in the supply lines, e.g. because of a leak in the system, the electric motors that drive the pumps are automatically switched off by hydraulic switches (*PSV*).

The undesirable pressure ripples that the pumps P_1 and P_2 might introduce into the system are smoothed out considerably by means of the 'hydraulic accumulators' (*A*). This ensures that the movement of the two carriages and the rotation of the main spindle are sufficiently uniform, which is vital to the quality of the work that can be done with COLATH. A non-return valve (*NRV*) is included in the line to *A*; this prevents pump reversal if a pump fails.

The problem of supplying oil from a stationary part of the machine to a moving one has been solved in a new way in COLATH (*fig. 8*). The final section of the supply pipe is supported in a hydrostatic bearing — hence with almost zero friction — in a sleeve *rigidly* fixed to the moving part of the machine. The moving sleeve is brought up to operating pressure via an aperture in the supply pipe — over the full travel of the moving part. The sleeve can therefore act as a source of supply for the moving part of the machine. At its other end the supply pipe is clamped flexibly (otherwise the structure might be deformed during the movement because of static over-determination).

The oil return

The oil used in motors H_1, H_2 and V is returned directly to the overflow reservoir R_o (fig. 7). The oil coming from the pressure-relief valves *PRV* also flows into R_o at first. The return oil from bearings *HB* passes through a coarse filter and also flows into R_o. The purpose of the filter is to remove any chips produced in the machining from the oil before it returns to R_o. The pump P_4 transfers the oil to the reservoir R via a fine filter and the cooler K. The various oil flows are continuously mixed in this reservoir, and the oil is then re-injected into the system.

The cooling system can be adjusted between two temperature limits; the amount of cooling does not depend on the temperature of the coolant.

Although the hydraulic system of COLATH bears a very close relationship to that of the PHM1000 [3] there are a number of essential differences, which will be examined in greater detail below. These include the modifications that were necessary for the change from manual operation to fast automatic operation and the improvements introduced as a result of the experience gained with the PHM1000.

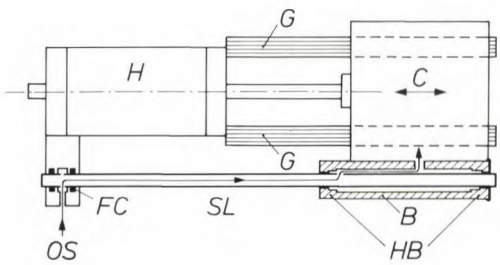


Fig. 8. The method of supplying oil from a stationary part to a moving one (carriage *C*) on the machine. *SL* supply pipe. *FC* flexible clamp (rubber). *B* sleeve, rigidly fixed to *C*; the length is approximately equal to the full distance of travel of *C*. The flow of the oil is indicated by the line *OS*. *HB* hydrostatic bearing (stepped). *G* guideway. *H* linear actuator (schematic).

The linear actuators and their control valves

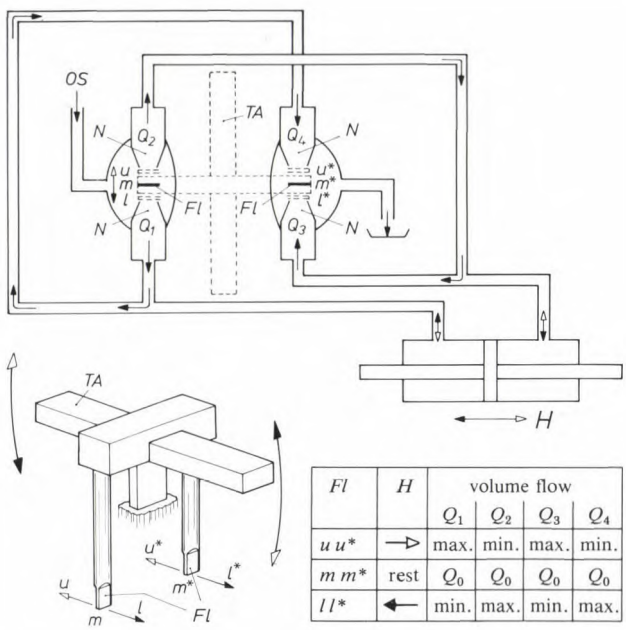
The motors H_1 and H_2 (fig. 7) are double-acting cylinders with a straight-through piston rod. In these symmetrical actuators the oil can be supplied to either side of the piston, depending on the position of the control valve. The valve then controls the actuators from *two* sides (instead of one) by means of four variable ports.

The control valve that we used (*fig. 9*) is a single-stage type; it contains two electromagnetically operated vanes that control the oil flow from or into the nozzles of four tubes, two at a time (the 'flapper-nozzle' construction for throttling the flow) [5].

The symmetrical design of the actuators, combined with the use of this control valve, has the advantage (compared with the PHM1000) that the stiffness of the oil column in the cylinders is effectively four times as great, so that the time constants of the control circuits are halved. The carriage can therefore be displaced more accurately and more quickly than in the PHM1000.

The stiffness of an oil column is defined as the ratio of a force acting on the oil to the resulting reduction in length (the compression) of the column; a practical unit is the newton per micron.

The coupling between piston rod and carriage has been designed to be self-adjusting. If the coupling were rigid, the structure would be statically over-determined and the linear displacement would be accompanied by bending or twisting. In the PHM1000 this was prevented by using special universal joints. In COLATH, however, with its fast control system, this cannot be done, because such a universal joint has a low stiffness. The principle of the new coupling is illustrated in *fig. 10*. To make the stiffness in the direc-



Fl	*H*	volume flow			
		Q_1	Q_2	Q_3	Q_4
*u u**	→	max.	min.	max.	min.
*m m**	rest	Q_0	Q_0	Q_0	Q_0
*l l**	←	min.	max.	min.	max.

Fig. 9. The control valve [5] for a linear actuator *H* (see fig. 7, actuators H_1 and H_2). *OS* oil supply. *N* nozzle. *Fl* flapper; *m,m** central position — motor is at rest; *u,u** and *l,l** extreme positions — motor is moving at maximum velocity to the right or left. The perspective drawing shows, in a somewhat simplified way, how the flappers are fixed to the moving armature *TA* of a torque motor (not illustrated). Q_1, ..., Q_4 volume flow of the oil.

[5] This control valve, Moog type 61, made by Moog Servocontrols (East Aurora, N.Y., U.S.A.), is the fastest known control valve. Its dynamic amplitude characteristic is almost flat up to about 400 Hz. The only disadvantage is that the maximum volume flow that the valve can deliver is fairly small, which limits the displacement velocities of the carriages.

tion of movement as high as possible, a hydrostatic bearing with a membrane double restrictor was incorporated for this direction [3]. In the directions at right angles to the direction of movement, however, the piston rod is not restricted and this means that small alignment errors are permissible.

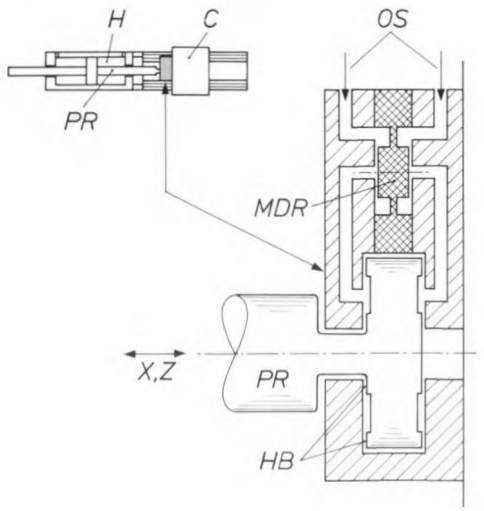


Fig. 10. The coupling between the piston rod *PR* of a linear actuator *H* and a carriage *C*. *MDR* membrane double restrictor; the effect of this disc-shaped element (cross-hatched) on the oil supply *OS* to the bearing *HB* is such that the stiffness in the directions of movement (*X*,*Z*) theoretically becomes infinitely great [3]. In the transverse directions the bearing has a stiffness of zero.

The rotary motor and its control valve

The motor *V* (fig. 7) is an improved version of the vane motor used in the PHM1000. The thrust bearing system is symmetrical with a greater distance between the bearings, an arrangement that gives a stiffer bearing system than in the PHM1000. The oil pressure in the motor itself (70 bars) is much higher than that in the bearing recesses (20 bars, half the working pressure); oil leakage resulting from this pressure difference — a source of undesirable effects on the speed of *V* — was reduced by small design modifications that made the leakage path in the motor longer and narrower. Because the motor and the main spindle are built as one integral structure, the modifications in the motor also led to changes in the design of the headstock (all of this in comparison with the PHM1000); this also offered an opportunity to increase the axial and radial stiffness of the main-spindle bearings, which improves the machining accuracy.

The control valve selected for *V* (fig. 7) is a two-stage design [6] whose first stage is almost identical to the control valve discussed earlier [5]. In the second stage there is hydraulic amplification of the control signal, so that the valve can supply a higher volume flow than the single-stage design.

Table II lists another set of data, including the stiffness of the various bearings and the lowest resonant frequency of the control valves and the hydraulic motors. This group of characteristic data for COLATH are closely related to the hydraulic system described here.

Table II. Some characteristic data for the hydraulic system of COLATH.

Movement	Bearing stiffness [c] (N/μm)		Resonant frequency (Hz)	
	vertical	horizontal	linear actuator [c]	control valve
X (toolholder carriage)	1200	600	218	400
Z (headstock carriage)	2400 [a]	1200 [a]	130	400
	radial	axial	rotary motor [c]	control valve
ϕ (rotation of main spindle) [b]	520	100	80	100

[a] Higher stiffness due to longer carriage.
[b] Tilting stiffness 100 N/μm, 50 mm away from centre-line of front bearing.
[c] Operating pressure for hydrostatic bearings 40 bars, for actuators and motor 70 bars (overpressures).

The laser measurement system

The control loops for the displacement of the two carriages (fig. 4) each contain a laser interferometer for determining the position of the carriage. *Fig. 11* shows a diagram of the entire measurement system for the two carriages; the block diagram relates to the carriage carrying the headstock.

The operation of such systems was explained in detail in a separate article in 1969 [7], when the PHM1000 lathe was also discussed in this journal. We shall therefore merely state the basic principles of the displacement measurements here.

The measurement of a displacement is based on the determination of the phase difference between two beams of light with almost identical frequencies (f_1 and f_2 in fig. 11*a*), which enter the measurement system polarized at right angles to each other, and are then allowed to interfere in a single direction via an interferometer (*I*) and an analyser (not shown). This gives an optical beat signal. The light modulation of the optical beat signal is transformed by a photodetector with amplifier (*R*) into an electrical measurement signal (*Meas* in fig. 11*a*) that contains the desired phase difference as a positive or negative Doppler effect. An electrical reference signal (*Ref* in fig. 11*a*)

[6] Moog, type 31.
[7] H. de Lang and G. Bouwhuis, Displacement measurement with a laser interferometer, Philips tech. Rev. **30**, 160-165, 1969. The laser measurement system used for COLATH has been put together from components made by Hewlett-Packard.

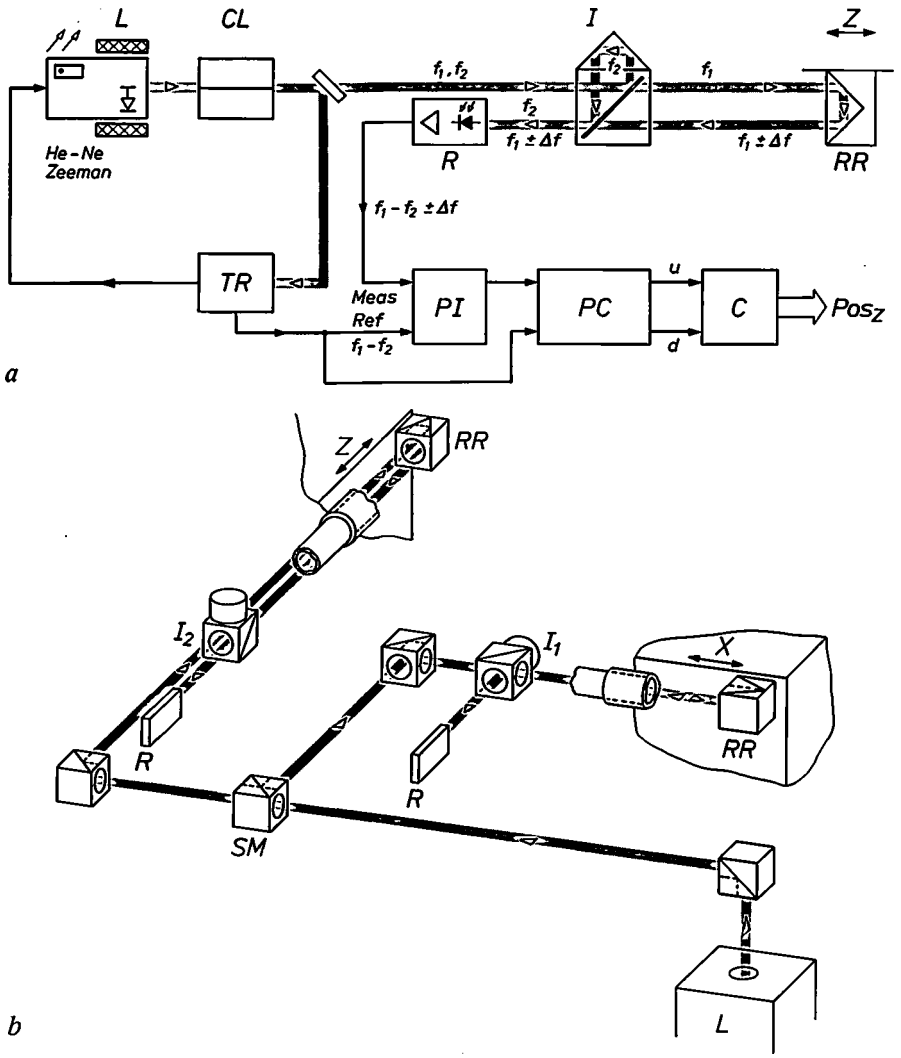


Fig. 11. The laser measurement system used in COLATH for measuring the carriage displacements (Z and X, see also fig. 3). *a*) Simplified block diagram of the laser measurement system with linear interferometer (double-beam) for the headstock carriage (Z). A single-beam interferometer is used for the toolholder carriage. L laser. *CL* collimator. I linear interferometer (double-beam). R receiver, consisting of photodetector and amplifier. *RR* retroreflector (cube-corner reflector). In the block diagram all the optical signals are indicated by light arrows and the electrical signals by dark arrows. Pos_Z position of carriage Z. *C* counter. *PC* pulse-conversion circuit. *PI* interpolation circuit. *Meas* measurement signal. *u* up: the pulse is added in *C*. *d* down: the pulse is subtracted in *C*. Δf frequency difference (due to Doppler effect). *TR* control circuits for stabilizing the laser frequencies (f_1 and f_2) and for generating the reference signal *Ref*. *b*) The perspective drawing shows the system with the two carriages. *SM* semi-transparent mirror (beam splitter). I_1 single-beam interferometer. I_2 linear interferometer (double-beam). The measurement beams are surrounded by telescopic tubes (partly shown).

without the Doppler effect is derived from the same two beams of light before the interferometer. The phase difference between the two electrical signals is continuously compared in the electronic measurement circuits from the start of the displacement. Whenever the displacement changes by an amount equal to $\lambda/4$ (λ is the wavelength of the laser light in the branch with the moving mirror *RR*) the measurement circuits generate a single pulse for a counter. The phase difference has then changed by 180 degrees. Depending on the sign of the change the pulse is added to, or subtracted from, the number of pulses already recorded. The counter therefore indicates the position of the carriage at any moment, as derived from a number of positive and negative displacements each of magnitude of $\lambda/4$.

The frequencies f_1 and f_2 are generated by a helium-neon (Zeeman-effect) laser [7]. In our case $f_1 \approx f_2 \approx 4.7 \times 10^{14}$ Hz; the absolute value of the difference is about 2.3×10^6 Hz. The behaviour of the measurement signal as a function of time can be described by an equation of the form $\sin[2\pi\{(f_1 - f_2) + 2\bar{v}/\lambda\}t]$, where $\bar{v}$ is the mean displacement velocity of the carriage during the duration t of the displacement. In the case of the reference signal the phase angle $4\pi\bar{v}t/\lambda$ does not appear in the sine argument. The term $2\bar{v}/\lambda$, which can be considered as a modification to the difference frequency $f_1 - f_2$, is the Doppler-effect term.

The resolution of the laser interferometer can be improved electronically by another factor of 10. This is done with the aid of an interpolation circuit (*PI*, fig. 11*a*). By applying frequency multiplication to the measurement signal and the reference signal the interpolator allows the smallest change in the reading of the pulse counter (one pulse) to be made to correspond to a displacement of one-tenth of the displacement equivalent to a phase difference of 180 degrees ($\lambda/4$). The resolution of the measurement system is therefore $\lambda/40$ and this is approximately equal to 16 nm.

The schematic diagram of the arrangement (fig. 11*b*) shows how the beam from the laser is split by means of a semi-transparent mirror (*SM*, beam splitter) into two beams at right angles to one another, each with half the intensity. The two measurement beams are enclosed in telescopic tubes between the interferometer and the moving retroreflector (or 'cube-corner mirror'), to give protection from the effect of local contamination and fluctuations in the density of the air on the beams. (Even the operator's breath might spoil the quality of the work, because of its higher temperature and higher content of carbon dioxide and water vapour, if the beams were not isolated from the environment.)

To make the system more compact, interferometers of a different type have been used; a *single*-beam interferometer for the toolholder carriage, which moves in the x-direction, and a *double*-beam interferometer for the carriage that moves in the z-direction.

The retroreflectors have the advantage that they make the interferometer insensitive to small unwanted rotary movements of the carriages. For displacement along a straight line it is impossible to eliminate rotation completely, however accurately the device has been made. The precise orientation of the retroreflector does not matter much in practice; the incident and reflected beams remain parallel and the total optical pathlength travelled does not change — and this is of course a great advantage when the measurement system is being set up.

The measurement unit

The wavelength of the laser light in the measurement beam is the standard on which the unit for the position measurements ($\lambda/40$) is based. To keep this standard as constant as possible we took three special precautions. First of all, the ambient air is carefully conditioned to stabilize its temperature to within ± 0.2 °C. Secondly, the oil in the hydraulic system circulates through the entire machine without interruption, while the water-cooling system keeps the temperature of the oil constant to within a few tenths of a degree. Thirdly, the machine is kept in operation continuously (even at night), and as a result there are no longer any significant variations in the temperatures of the various parts of the machine.

Our subsequent experience with COLATH suggests that these three precautions are sufficient, and this seems reasonable since the final machining of a product only takes a few minutes. Any changes in the ambient conditions (moisture content, temperature, pressure) that might have serious consequences *cannot* arise in such a short time.

A single control pulse is nominally equivalent to a dimensional unit equal to 1/64 μm, as we shall see in the section on numerical control. This measurement unit is a little smaller than $\lambda/40$, the unit corresponding to a single pulse of the measurement signal in the feedback control loop of the machine. To ensure that any desired number of control pulses M still results in a workpiece dimension of $M \times 1/64$ μm, the control program must reduce the number M slightly. This is achieved by continuously multiplying the control signals by the scale factor 0.98764300. This scale factor can be determined by a calibration, that is to say by measuring displacements on the machine directly with an independent and correctly calibrated laser interferometer. A better method, but not such a fast one, is to proceed as follows. A desired workpiece dimension (a_0) is selected more or less at random. This dimension a_0 divided by 1/64 μm gives a number of control pulses (M_0). M_0 control pulses are then fed to the control system of the machine, so that a workpiece, the 'test piece', is produced. If, finally, the actual dimension of the test piece is measured — *completely independently* of COLATH — and this is then divided into the dimension a_0, then the desired scale factor is obtained. It is of course also possible to carry out this calibration more than once and for the ultimate in precision the scale factor is redetermined a number of times.

Numerical control

The main element of the control unit (fig. 1) is a minicomputer (Philips P9201). The numerical-control system also comprises electronic interface circuits and two control panels, one on the lathe itself and the other on a control cabinet. The storage capacity of the computer is approximately 8000 words of 16 bits and the cycle time is 0.96 μs. This control system is particularly versatile; changes usually only require modifications to the software for the computer. The most important advantage of this is that COLATH can still be used in the usual way for machining work in the original unchanged program while modifications are being carried out.

In the software for COLATH a clear distinction is made between the P9201 control program relating to the lathe itself and the 'part program'. A part (or workpiece) program of this kind is prepared 'externally', that is to say on a separate computer and independently of the control unit (we shall return to this later). It describes only the movements of the lathe (displacements and speeds) that are necessary to make

the particular workpiece, and it is used, in the form of a punched tape, as input for the control program.

The control panel attached to the lathe is mainly for the operations where a direct view of the workpiece is desirable; for example positioning the tool with respect to the workpiece at the start of a new program.

The panel on the control cabinet houses the display units for the counters (fig. 11) that indicate the positions of the two carriages. A separate display unit can be used to indicate a third machine function, which can be either the speed of the main spindle, the contour velocity of the tool, or the 'block number', which indicates the part (the program block) being executed on the punched tape containing the workpiece program.

The speed of the main spindle can be manually controlled from either panel.

The P9201 control program

The control program ensures that all the instructions that the COLATH operator gives to the computer are carried out. The heart of the control program is the 'scheduler', which causes the various combinations of the subprograms contained in the computer to be started and finished in their various priorities. This, of course, requires a communication system that continuously records and releases control data. In our case use is made of periodic sampling (sampling rate 500 Hz); as a result of this the data obtained from the lathe is renewed every 2 ms and new control data is fed to the lathe at the same rate. In the intervals in between, i.e. preceding each sampling, certain preparations are taking place, such as starting the available tape reader (again), processing the collected data into useful new information, etc.

The control program is designed in such a way that only straight-line movements are performed, and this is fully compatible with the linear approximation method used in the part programming (we shall return to this in the next section). We aim to extend the control program in due course by adding a facility for nonlinear interpolation, which will approximate the desired contours by curved line elements. This kind of interpolation can lead to better optical performance, e.g. because reflection losses are more uniformly distributed.

One of the important functions of the control program is to monitor the contour velocity of the tool.

The contour velocity of the tool is the velocity at which the tip of the tool follows the desired contour in the material being turned. In the case of COLATH the contour is a curve in the x,z-plane. The contour velocity is the vector sum of the displacement velocity of the toolholder carriage and a velocity equal and opposite to that of the headstock carriage (fig. 3).

Within each program block the control program selects the velocity in such a way that the deviation in the final position for the block is as small as possible. The punched tape mentioned before only gives the desired contour velocity. If for example there is a slight kink in the path to be followed, the control program adjusts the deceleration to a lower value of the velocity and adjusts the subsequent acceleration back to the desired contour velocity. The monitoring of the velocity also ensures that in all circumstances, including changing from one program block to another, the velocity remains between the minimum and maximum permissible values, and that the movements take place as smoothly as possible.

The computation unit in the control program has been taken as 1/128 μm. This was done to allow for any inaccuracies in the computation and to permit the accuracy of the lathe to be improved at some future time if so desired. The computer would then have to be replaced. In the present situation the control-output unit corresponds to 1/64 μm, with the result that all calculated values have to be halved.

Part programming

A part program is processed on a Philips P856 minicomputer, completely independently of any work being executed on COLATH; this results in a punched tape with numerical data for the workpiece. The processing is mostly done by using special routines written in FORTRAN, or sometimes with the aid of the programming language MICROAPT, derived from the APT language [8]. The FORTRAN routines have been prepared mainly for making batches of similar workpieces, i.e. a 'workpiece family'. There is for example a 'universal' program that can be used for generating the data for all possible aspherical contours.

When the punched tape is being prepared, the main objective is to compute the tool path, the contour that the diamond tip has to cut in the material of the workpiece. However, this computation process can only approximate the theoretically desired contour of the workpieces to be turned, and it does this by means of straight lines (linear interpolation); the accuracy of this approximation depends on the total width of the 'tolerance zone' (*fig. 12*). This value, for example 0.1 μm for rough-turning a workpiece and 0.01 μm for finish-turning, is fed to the P856 beforehand; this means that the P856 continues to calculate the path of the tool until the path obtained is inside the tolerance zone everywhere. The path data obtained, the x- and z-coordinates of the starting and end points of each

[8] J. Vlietstra, The APT programming language for the numerical control of machine tools, Philips tech. Rev. 28, 329-335, 1967.

line element, as well as a number of subsidiary functions (contour velocity, stop codes, etc.) are entered on the punched tape. The punched tape is subdivided into the program blocks mentioned earlier. Each block contains the information for one line element. For a complicated contour with small tolerances the number of program blocks can therefore be quite high. A plotter connected to the P856 is used to give a verification of the punched tape. The tape is then fed to the control unit of the numerical-control system (the P9201) at a speed of 300 characters per second via an optical 8 channel reader, in portions of several blocks, while the control program executes the work instruction block by block.

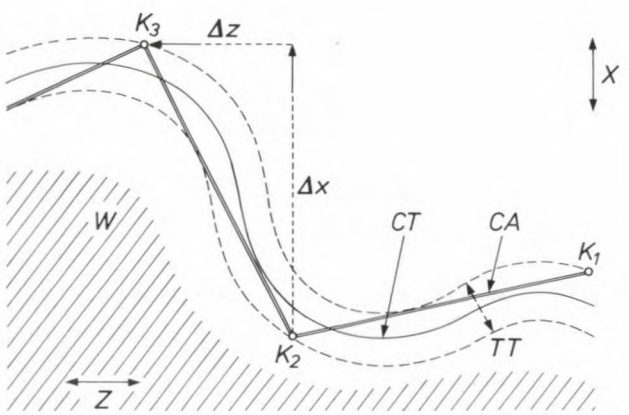


Fig. 12. Example of the linear approximation method in part programming. *CT* the desired contour, e.g. an ideal optical contour. *CA*, with points K_1, K_2, K_3, etc., is the approximation to *CT* by the computing process. *TT* tolerance zone. The computing process is designed so as to minimize the number of points K_i for any contour to be machined. The Δx and Δz shown apply for an imaginary diamond of zero radius. The two increments are modified by the computing process to correct for the non-zero radius of the diamond. They then form the main numerical workpiece data for one program block on the punched tape. The punched tape — produced completely independently of COLATH — serves as the input for the control program. *W* workpiece. *X*, *Z* movements of the toolholder and headstock carriages.

Control loops

Linear movements

The displacements of the two carriages, in the *x*- and *z*-directions, are controlled almost identically. The basic principle of the control procedure has already been discussed in the section on the general

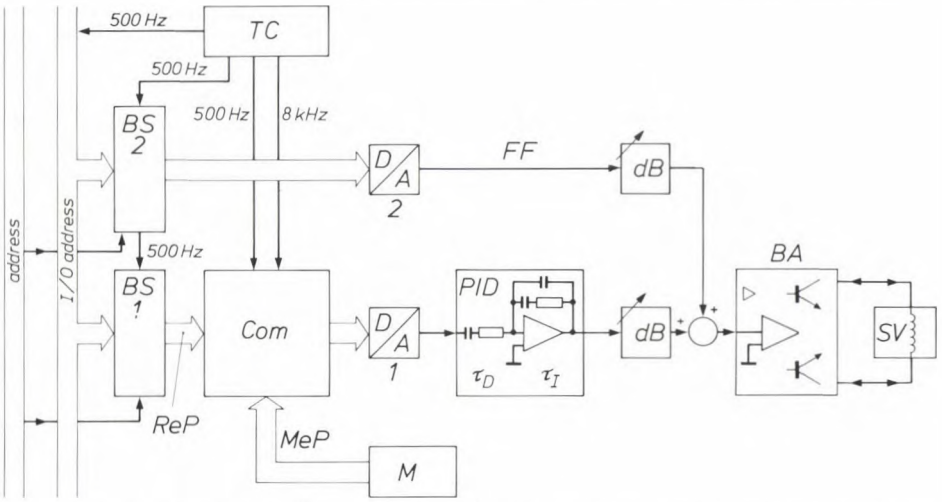


Fig. 13. Part of the block diagram for the control of the displacement of a carriage (see also fig. 4). *TC* clock-pulse generator. *BS* buffer store. *Com* comparator. *dB* variable attenuator. *BA* balanced output amplifier. The remaining symbols are explained in the caption to fig. 4.

features of the new lathe, and a block diagram has been given (fig. 4). Some more detailed features, mainly elements in the forward path of the control loop, are illustrated in *fig. 13*.

Updated data on the *desired* displacement are stored 500 times per second in a buffer register (*BS1*) in the form of a 12 bit word. A change in the least-significant bit results in a displacement of $\lambda/40$. The contents of *BS1* are shifted directly to the comparator (*Com*) where the binary data on the *measured* displacement, again in the form of a 12 bit word, is also arriving continuously.

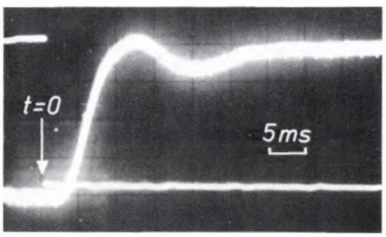


Fig. 14. Oscillogram of the step response of the control system (see also fig. 4) for displacements of the toolholder carriage. The desired displacement of 1 μm in the *x*-direction is achieved within about 10 ms; there is a slight overshoot, which has been reduced to zero after about one-and-a-half cycles.

The desired and measured positions are compared 16 times in each 2 ms sampling period; the binary difference signal goes to the input of the digital-to-analog converter *D/A 1*.

The step instruction being made at the time of sampling — a desired carriage displacement, is divided by means of an interpolator into 16 equal sub-steps in the ensuing period of 2 ms. This prevents the control loop from having to react to too large a difference signal at any sampling time.

Rotary movement

The speed of the main spindle (fig. 3) is continuously variable — up to a maximum of about 1000 revolutions per minute — by instructions from the control computer. Feedback from a tachogenerator ensures that the speed does not change. Our experience shows that a simple control system of this kind is adequate provided that machining is restricted to the turning of plastics and non-ferrous metals with a diamond-tipped tool. In these applications the quality of the work is not affected by the change in the speed that occurs as a function of the diameter at which the workpiece is being machined.

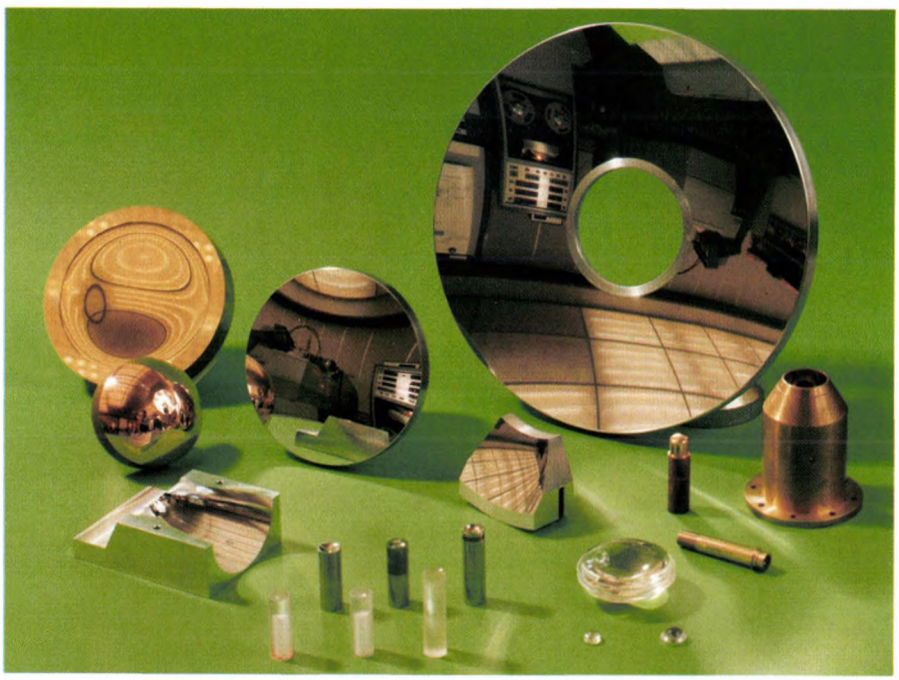


Fig. 15. Some optical components (and a few others) that have been made on COLATH.

The analog output signal from *D/A 1*, a voltage between -6 and $+6$ V, is fed via the electronic controller *PID*, an attenuator, and the output amplifier *BA* to the hydraulic control valve *SV*. This valve controls the linear actuator (H_1, H_2 in fig. 7) and the carriage connected to it. The controller contains an integrating circuit and a differentiating circuit, each with an adjustable time constant (*RC* time constants τ_I and τ_D). This feature allows the behaviour of the control loop to be improved. To obtain the best possible improvement — which amongst other things implies that the difference between the desired and actual positions is small enough — a parallel branch has been added to the forward path of the control loop. This parallel branch supplies a signal to the output amplifier via *BS2*, *D/A 2*, and an attenuator — a signal that is proportional to the velocity $\dot{x}$ or $\dot{z}$ (feedforward).

Fig. 14 illustrates an example of the step response of the entire system with closed-loop control. The dynamic behaviour, i.e. the response rate and the overshoot, is entirely acceptable.

Workpieces

Fig. 15 illustrates a collection of objects that have been made with the new lathe. The most important work done so far has been the making of bi-aspheric lenses, single lenses in which the two refracting surfaces differ from a true sphere in an extremely precisely specified manner. The bi-aspheric lenses can be characterized as high-grade 'diffraction-limited' optical elements, with additional advantages such as small size and mass. A design for such a bi-aspheric lens is illustrated in *fig. 16*. This lens is suitable for reading out information in a 'VLP' player [9] that has

[9] J. Haisma, E. Hugues and C. Babolat, Realization of a bi-aspherical objective lens for the Philips Video Long Play system, Optics Letters **4**, 70-72, 1979.

a helium-neon laser as the light source (wavelength 632.8 nm). *Table III* summarizes the data for the desired aspherical contours. A batch of 25 lenses based on this design have already been machined one after another on COLATH.

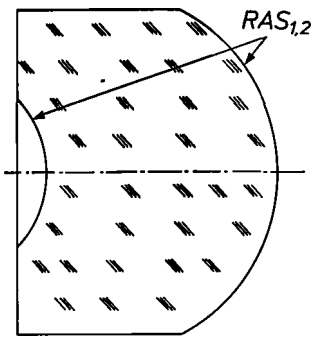


Fig. 16. A design for a bi-aspheric lens, as turned on COLATH. This lens with two aspherical refracting surfaces ($RAS_{1,2}$) is suitable for 'VLP' players with a helium-neon laser as the light source. The form error should be no more than 0.2 μm, and the surface roughness no more than about 20 nm (peak-to-valley). The diameter of the lens is 8 mm, the maximum length 6.5 ± 0.01 mm.

Table III. The calculated coefficients of the contour function $z(x)$ for the refracting surfaces of a bi-aspheric lens that can be used for a 'VLP' player. Lens material: polymethyl methacrylate. Light source: helium-neon laser, wavelength 632.8 nm.

Coefficient	Contour function $z = ax^2/[1 + \{1 - (1 + b)a^2x^2\}^{\frac{1}{2}}] + \sum_{i=2}^{10} c_i x^{2i}$ (z and x in mm)	
	surface *1*	surface *2*
	$n = 1.4908$ at $\lambda = 632.8$ nm	
a	$+0.240000$	-0.51506×10^{-1}
b	-0.29461985	-0.79999995
c_2	$-0.23658267 \times 10^{-3}$	$+0.42097500 \times 10^{-2}$
c_3	$+0.18060664 \times 10^{-5}$	$-0.81530879 \times 10^{-4}$
c_4	$-0.38227121 \times 10^{-5}$	$-0.60478503 \times 10^{-5}$
c_5	$+0.14036973 \times 10^{-5}$	$+0.40506688 \times 10^{-4}$
c_6	$-0.29596792 \times 10^{-6}$	$-0.31471338 \times 10^{-4}$
c_7	$+0.37374523 \times 10^{-7}$	$+0.13752718 \times 10^{-4}$
c_8	$-0.27752954 \times 10^{-8}$	$-0.34047096 \times 10^{-5}$
c_9	$+0.11171180 \times 10^{-9}$	$+0.44607171 \times 10^{-6}$
c_{10}	$-0.18717534 \times 10^{-11}$	$-0.23890333 \times 10^{-7}$

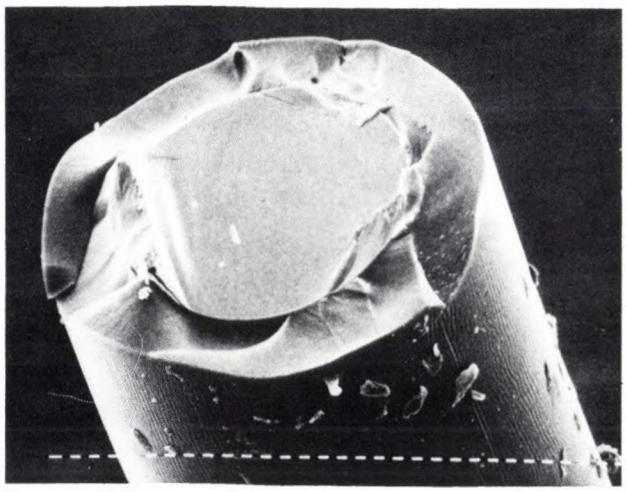
Interferometric measurements of the quality of these lenses have shown that the inaccuracy of each refractive surface is only 0.1 to 0.2 μm. One of the lenses has been mounted in a 'VLP' player to find out if the picture information could be read out without annoying errors. The results for both stationary and moving pictures were just as good as those obtained with a 'conventional' read-out system containing no less than four spherical lenses.

In the meantime COLATH has also been used to produce a number of components for use in the optical recording of X-ray images, e.g. in X-ray diagnostics. These transparent elements, called 'correction plates', have a flat back and an aspherical front, i.e. they are true correction plates (and not lenses) for certain optical image errors. The geometrical tolerance need not be smaller than 0.5 μm; the diameter of the plates we have produced is between 30 and 70 mm.

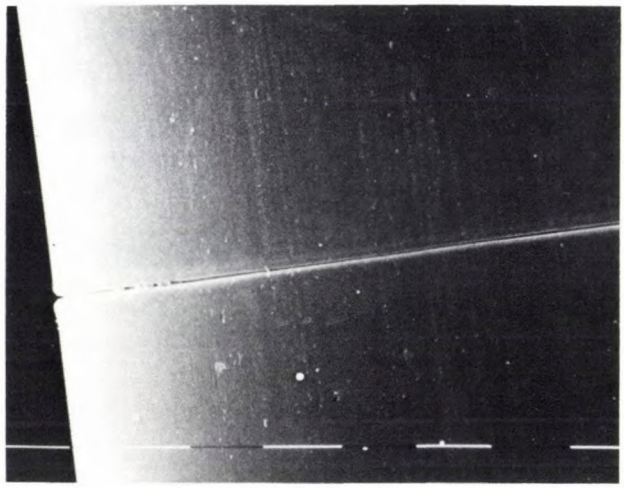
An optical system for X-ray diagnostic equipment usually has three objective lenses: a basic objective that in principle forms an image at infinity of the electronically intensified X-ray picture — for individual observation, a television objective that produces an image of the X-ray pictures on the sensitive layer of a television camera, and a photographic objective that forms an image so that the X-ray picture can be stored on film. By incorporating aspherical correction plates the size of these lenses can be kept to a minimum, whereas the distortion and the field curvature are still sufficiently small for the images to remain completely acceptable. One problem is that the machining marks cause extra scattering of the light, reducing the contrast of the image. Studies are continuing to find out how to minimize this scatter.

COLATH has also been used successfully for making metal mirrors and germanium lenses for infrared equipment. Here the geometrical precision does not usually have to be much better than 1 μm.

Summary. Description of the construction and characteristics of a numerically controlled precision lathe, COLATH. With a diamond-tipped tool it can be used to machine surfaces of revolution of 'optically smooth' (diffraction-limited) quality in non-ferrous metals and alloys and in plastics. Maximum workpiece dimensions are 200 mm in diameter (2× maximum movement of toolholder carriage) and 200 mm in length (maximum movement of headstock carriage); form error is < 0.5 μm, surface roughness (peak-to-valley) is about 20 nm; maximum velocity is 180 mm/min (both carriages), main-spindle speed 0-1000 rev/min. A minicomputer (P9201) connected to a 2-axis laser measurement system (resolution 16 nm) controls the displacements of the carriages by means of hydraulic servomotors. The main spindle is driven by a hydraulic vane motor, which is integral with it. All the bearings are hydrostatic (with oil as the working fluid). The temperatures of the ambient air and the oil in the hydraulic system are stabilized. The measurement beams of the laser are enclosed in telescopic tubes. The software consists of the P9201 control program and externally produced part programs on punched tape. In principle, modifications in the control system only require changes to be made in the P9201 control program. A universal part program ('family program') is available for aspheric lenses. The lathe has been in use since 1978, and aspheric lenses have now been made for infrared and visible light.



1



2

Precision fracture of optical glass fibres

In the practical application of optical glass fibres in telecommunication it is necessary to be able to break the fibres to give a smooth fracture surface accurately perpendicular to the axis. In the methods that have so far been used the fibre is nicked on one side, by scratching it with a diamond [1] or by heating it with a hot filament [2], and the two ends are then pulled apart. It is found that in these methods the fracture face is by no means always smooth and is generally not truly perpendicular to the axis of the fibre, as *fig. 1* shows for a fibre, with a plastic cladding, that has been nicked by a diamond. In a new method [3], in which these inaccuracies occur much less frequently, the problem is tackled in the following way. A scratch is made right round the fibre with a diamond cutter, and the two ends are then pulled apart. This method can be used for all kinds of fibres: soft glass or fused silica, with a cladding of plastic or glass (usually protected by a plastic coating).

To make the scratch, the cutter, which rotates around the tensioned fibre, is first brought up to speed and slowly moved towards the fibre and pressed against it. The cutter is then pressed against the fibre with the appropriate force. A perfect scratch is obtained in this way, as *fig. 2* shows (the plastic protective coating has been removed here). The values for the force and speed for such a result follow from a study of the grinding of brittle materials [4]. The fracture produced on pulling makes an angle of less than 2° with the normal to the fibre axis and has a perfectly smooth surface. *Fig. 3* shows this for a completely glass fibre (diameter 100 μm) with a plastic coating, and *fig. 4* shows the fracture for a fibre with plastic cladding (total diameter 300 μm). In fig. 4 the track made by the cutter, which has gone through the cladding in a single revolution, is clearly evident.

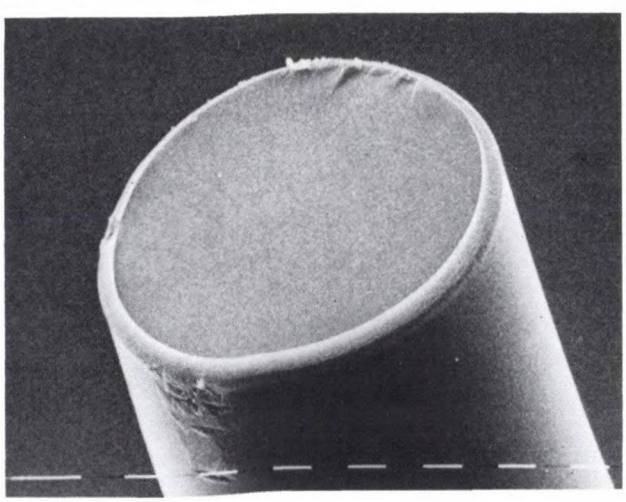
The design and construction of the cutting device that gave the results shown here and the studies that led to its construction were carried out by a team consisting of C. J. G. Verwer, Ing. J. C. G. Teunissen, A. J. Luyten, Ing. J. A. Luijendijk, Ing. G. Kuyt, Ir G. D. Khoe, A. J. J. Franken and Ing. R. Brehm, all with Philips Research Laboratories, Eindhoven.

[1] D. Gloge, P. W. Smith, D. L. Bisbee and E. L. Chinnock, Bell Syst. tech. J. **52**, 1579, 1973.

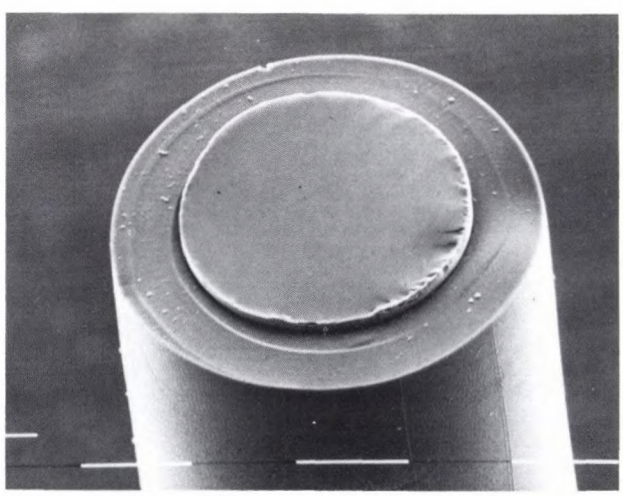
[2] W. J. J. van Hoppe, G. D. Khoe, G. Kuyt and H. F. G. Smulders, Philips tech. Rev. **37**, 89, 1977.

[3] G. D. Khoe, G. Kuyt and J. A. Luijendijk, Appl. Optics **20**, 707, 1981.

[4] A. Broese van Groenou and J. D. B. Veldkamp, Philips tech. Rev. **38**, 105, 1978/79.



3



4

Application of cryopumps in industrial vacuum technology

J. J. Scheer and J. Visser

'Nature abhors a vacuum', Aristotle is supposed to have said. Even today it is no great exaggeration to say that vacuum technologists greet the industrial application of a new type of vacuum pump with some trepidation. However, the cryopump has been accepted fairly rapidly as a means of reaching a low vacuum in a short time, with extremely low contamination from the residual gases in the vacuum chamber. The applications so far have remained limited to certain areas of industrial vacuum technology. By adapting the design to the intended use in each case the authors have used the cryopump successfully in widely differing pressure ranges, from rough vacuum to well into the ultra-high-vacuum range. Four applications are described; in two of them cryopumps have never been used before.

Introduction

For many years the diffusion pump, invented by W. Gaede at the beginning of the century, was the only pump that the vacuum technologist could use for reaching the low pressures required in the manufacture of thermionic devices, the operation of electron microscopes and cyclotrons, the deposition of thin films and other applications.

As is well known, in a diffusion pump use is made of a liquid, the pump fluid — a mineral or synthetic oil or mercury — and consequently the vacuum system is always contaminated by the saturated vapour of the pump fluid at the temperature in the neighbourhood of the pump aperture. There is also 'backstreaming' of pump-fluid molecules, which condense on the walls of the vacuum system and any objects inside it.

Although the contamination of the residual gases in the vacuum system by vapour from the pump fluid can be reduced to some extent by using water-cooled baffles — or preferably with liquid-nitrogen-cooled baffles — there was still a need for pumps in which there was no direct contact with a liquid. This was because the requirements for the final pressure and especially for the permissible degree of contamination of the residual gas became more and more demanding.

In the sixties, therefore, a great deal of interest in the turbomolecular pump began to arise. In this pump the gas molecules — originally moving in all directions — have a preferential direction imposed upon them by collision with the rapidly rotating rotor blades. Since there is no direct contact with a fluid, there is less contamination of the residual gas in the vacuum system.

The diffusion pump and the turbomolecular pump both belong to the group known as 'transport pumps', a kind of pump that removes the gas from the vacuum system and transfers it to the backing pump. This pump, filled with mineral oil, compresses the gas to atmospheric pressure. The backing pump itself is a source of contamination by hydrocarbons because oil molecules can diffuse into the vacuum system in the opposite direction to the general flow. This happens mainly when the pressure in the backing system is low enough for the flow to be molecular rather than laminar.

A complete absence of hydrocarbons can be attained by using the cryopump. In this kind of pump the gas to be pumped is not removed from the vacuum system but attached — without any intervention from a liquid — to a surface in the vacuum system. A fortuitous advantage in cryopumping is that the backing pump only has to be run briefly at the start of the pumping run. In addition to the absence of a contaminating liquid the cryopump also has the advantage that high pumping speeds can be reached, and more easily than with diffusion or turbomolecular pumps.

We shall now look first at the operation of cryopumps, and with the aid of an example we shall indicate the factors that have to be taken into account in designing cryopumps. Next we shall describe four cases in which we have successfully employed cryopumps. Applications in pumping a gas-analysis system and in ultra-high-vacuum technology are — as far as we know — ones in which other pump types or meth-

Dr J. J. Scheer and J. Visser are with Philips Research Laboratories, Eindhoven.

ods have always been used until now. While applications in the pumping of equipment for the sputtering and evaporation (vacuum deposition) of thin films are not new, they clearly show that a good understanding of the operation of these pumps is necessary in producing an appropriate design and hence the best results. These applications will be described in order of decreasing pressure.

The operation of a cryopump

As described in an earlier article [1] in this journal, the operation of cryopumps depends on two separate physical effects. One effect is condensation of the gas on the cooled surface; the pumping effect occurs here if the operating pressure is higher than the saturated vapour pressure at the temperature of the surface. The other effect is physical adsorption of the gas molecules; the cooling of the surface increases the 'time of sojourn' of the molecules on the surface. In the first case relatively thick layers of gas are condensed on the surface and then 'frozen in'; in the second there are one or more monomolecular layers on the surface. The pumping effect due to condensation stops when the thermodynamic equilibrium between the solid material and its saturated vapour pressure has been reached; the pumping effect due to adsorption stops when all the available sites on the surface have become occupied by atoms from the gas, with the gas atoms and the atoms of the wall attached to one another by polar forces.

We shall look at condensation first, and we should therefore consider the variation of the saturated vapour pressure of the gases commonly present in a vacuum system as a function of the absolute temperature; see *fig. 1*. This shows that with a cold panel cooled by liquid nitrogen (boiling point 77 K), for example, H_2O can be pumped readily. The vapour pressure of CO_2 at this temperature, however, is still about 10^{-5} Pa, and N_2 and O_2 cannot be pumped with such a panel.

If liquid helium (boiling point 4.2 K) is used, however, practically all gases can be pumped by condensation, except H_2 (residual pressure about 10^{-4} Pa) and He itself. With commercially available refrigerating machines a panel temperature of 20 K can be reached, and more easily than by using liquid helium. All gases except neon, hydrogen and helium can be condensed adequately with such a system. Since hydrogen is a gas that is formed in virtually all industrial processes, it is clear that in most cases condensation alone will not be effective.

We shall now look more closely at the adsorption effect. In adsorption the gas molecules are only attached to the surface for a certain time. The time of sojourn of the gas molecules on the wall depends on the wall temperature and the combination of wall material and gas. At low temperature and high heat of adsorption (the heat liberated when a quantity of gas molecules is adsorbed) the time of sojourn (we shall say more about this later) becomes larger. It will be clear that the maximum quantity of gas that can be taken up by adsorption is determined by the actual surface area of the panel. It is desirable to use a wall material of high porosity. Activated charcoal is found to be very suitable; it has a surface area of nearly 900 m^2 per gram of material. One gram of carbon can capture a maximum of 30 000 Pa·l of N_2, for example (in the adsorption of a monolayer). Zeolites are also found to have a high adsorptivity for gas molecules. Carbon, however, does not enter into chemical reaction with the adsorbed gas molecules (zeolite reacts with adsorbed H_2O), and so the adsorbed gas is readily released at room temperature (regeneration). The maximum quantity of gas that can be taken up by adsorption is determined by the quantity of activated charcoal applied to the panel. The quantity of gas that can be taken up by condensation is virtually unlimited, however.

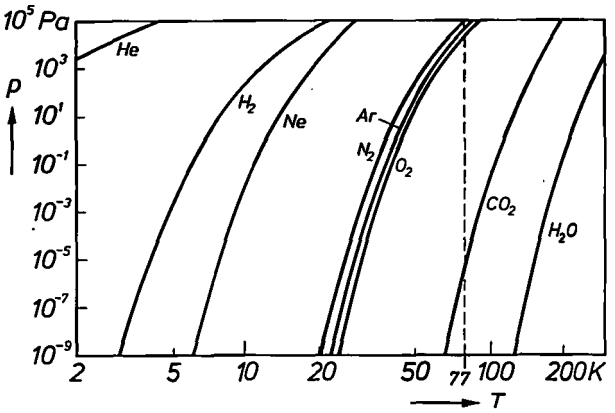


Fig. 1. The saturation vapour pressure p as a function of the absolute temperature T, for the gases helium, hydrogen, neon, nitrogen, argon, oxygen, carbon dioxide and water. The points of intersection of the appropriate curves with the vertical dashed line indicate the final pressures to which Ar, O_2 and CO_2 can be pumped with a cold panel cooled by liquid nitrogen. The intersection with the curve for H_2O is outside the figure.

We shall now look at the theoretical pumping speed that can be obtained with cryopumps. This is given by the relation

$$\frac{dN}{dt} = cvA_p - \frac{NA_p}{\tau A} \text{ molecules/s}, \qquad (1)$$

where N is the number of molecules 'trapped' (adsorbed or condensed) by the surface. The first term on the right-hand side represents the number of molecules that strike the surface in unit time, the second

[1] A. Venema, Philips tech. Rev. 28, 355, 1967.

term the number that leave it in unit time. The quantity c can be taken as the 'sticking probability', the probability that a molecule that strikes the surface does not immediately leave it again. A_p represents the projected surface of the cold panel and A represents its actual microscopic surface; both A_p and A are measured in m^2. As the relation shows, A_p determines the value of the pumping speed; A determines the maximum quantity of gas that can be taken up. The factor v is equal to the number of molecules striking unit area of the wall in unit time, and is given by the relation

$$v = 2.6 \times 10^{24} \frac{p}{\sqrt{MT}} \text{ molecules/m}^2\,\text{s}, \qquad (2)$$

where M is equal to the molecular weight in kg per kmol, T is the absolute temperature and p is the pressure in Pa of the gas being pumped. The quantity τ in (1) is equal to the mean time of sojourn of the molecules on the wall, and is given by

$$\tau = \tau_0 \exp(E_a/RT_w), \qquad (3)$$

where $\tau_0 \approx 10^{-13}$ s, the period of vibration of the gas molecules, E_a is the adsorption energy of the gas molecules at the surface and T_w is the temperature of the surface. It follows from (3) that cooling the wall makes the second term of the right-hand side of (1) smaller. If the temperature of the wall is made equal to absolute zero, this term becomes zero: all the molecules that strike the wall stay there.

In practice c is found to approach the value 1 very closely [2], particularly for condensation. If we assume that at the start of the pumping run no gas molecules have yet been adsorbed at the wall (N is virtually equal to zero, the pump has been regenerated) or that the wall temperature approaches absolute zero, then making use of (1) and (2), the pumping speed is given by

$$S = \frac{dN}{dt} = 2.6 \times 10^{24} \frac{pA_p}{\sqrt{MT}} \text{ molecules/s}$$

or (4)

$$S \approx 10^7 \frac{pA_p}{\sqrt{MT}} \text{ Pa} \cdot \text{l/s}.$$

This relation is similar to the relation for the theoretical maximum pumping speed of an aperture through which all the incident gas molecules can disappear. The relation for the theoretical pumping speed of diffusion pumps also takes this form. In both types of pump only a fraction (30 to 70%) of the theoretical pumping speed can be attained because of the presence of flow resistances such as baffles and valves.

From the heat of condensation and the heat that must be removed from the gas to cool it from room temperature to the temperature of the cold panels we can find the refrigerating power required for the generator that has to keep the panels at their low temperature. For the gases most commonly encountered a refrigerating power of 1 W is adequate for a gas flow of 250 Pa·l/s.

If the pumping speed decreases because of saturation of the layer of activated charcoal, e.g. because it has adsorbed large quantities of H_2, a cryopump should be regenerated. This can easily be done by bringing the pump to room temperature for several hours and pumping away the liberated gas with the backing pump.

Some design aspects

We shall now illustrate some of the design aspects that have to be taken into account, by referring to one of our designs for a cryopump, intended for pumping a system in which a great deal of H_2 is evolved in the presence of large amounts of gases that condense easily.

Although there are other methods of producing cold, in the cryopump designs described here we always used the Philips K20 cryogenerator, type PW 7600 [3]. This machine (see *fig. 2*) is based on the Stirling cycle [4], and, since the process takes place in

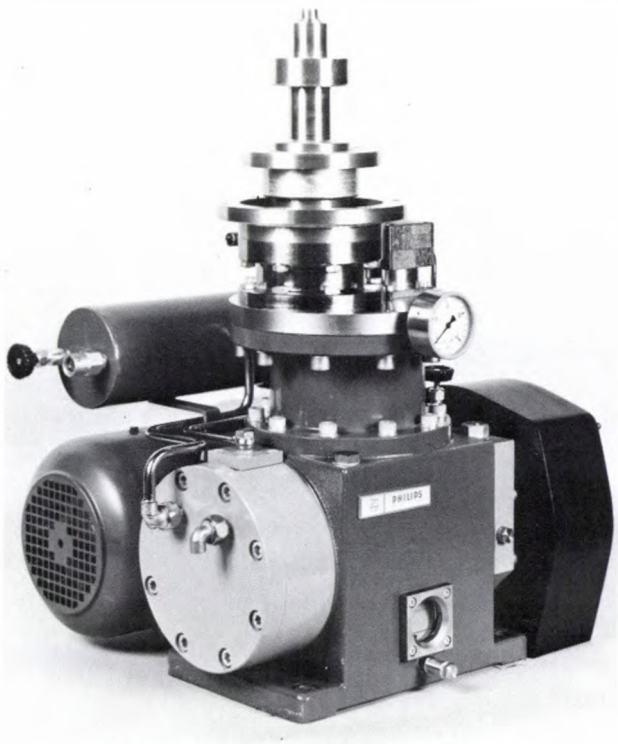


Fig. 2. The Philips K20 cryogenerator (type number PW 7600), operating in the Stirling cycle. The gauge indicates the pressure of the helium used in the machine as the working gas.

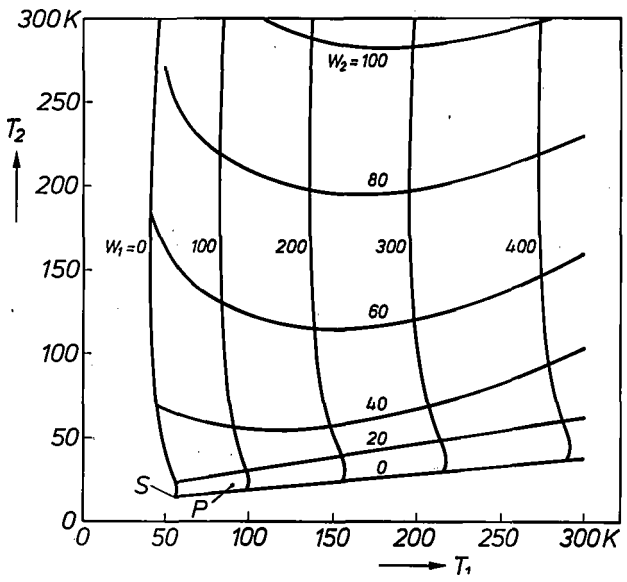


Fig. 3. Characteristic curves for the K20 cryogenerator. T_1, T_2 temperatures of the first and second stages in degrees absolute. The approximately horizontal curves give the various combinations of T_2 and T_1 that are obtained with constant refrigerating power W_2 for the second stage and varying refrigerating power W_1 for the first stage. The same applies for the approximately vertical curves with constant W_1 and varying W_2. The operating point of the machine is determined by the refrigerating powers W_1 and W_2 required at the two stages. If the operating point is known, the associated temperatures T_1 and T_2 can be determined. The point P is approximately the operating point for the application of the machine as a cryopump. The point S gives the theoretically attainable minimum temperatures under no-load conditions: 14 K and 60 K.

two stages, it can deliver cold simultaneously in two temperature ranges. The relation between the refrigerating powers required at the first and second stages and the associated temperatures can be derived from *fig. 3*, which shows the characteristic curves for the machine. Since the curves do not differ greatly from a straight line, the temperature of one of the stages does not vary very much if the power taken from the other stage is varied. The point P in the figure is the operating point for the application described here. At the first stage 80 W of cold is produced at 80 K, and at the second stage 10 W of cold at 20 K.

Figs 4 and *5* show how the cold panels are arranged [5]. At the outside there is first of all a baffle A, connected to the first stage of the machine and at a temperature of 80 K. The main function of this baffle is to give protection from the thermal radiation from the relatively warm environment. The set of chevron-shaped baffles B are also at a temperature of 80 K. They pump the readily condensed gases such as H_2O. The sets of panels C and D are connected to the second stage of the machine and have a temperature of 20 K. The panels C are also chevron-shaped, and the gases N_2, O_2 and CO_2 condense on them. The panels D are mounted radially and are coated with a layer of activated charcoal. The panels C shield the panels D in such a way that oxygen and nitrogen do not block the pores of the charcoal. The gases that can still reach the carbon-coated panels are H_2, He and Ne; hydrogen is the gas most commonly present.

The radial arrangement of the carbon-coated panels is one of the ways of reaching a compromise between the largest possible carbon-coated surface and easy access to the panels for gas molecules. It is found that H_2 molecules that are adsorbed in the sorption material at 20 K are mainly taken up by the part of the surface where the gas molecules are first incident. Saturation of that part of the sorption material causes the pumping speed to fall too soon. The radial arrangement gives a reasonably uniform distribution of the gas molecules over the surface.

We have carried out an investigation to find the optimum grain size for the layer of activated charcoal.

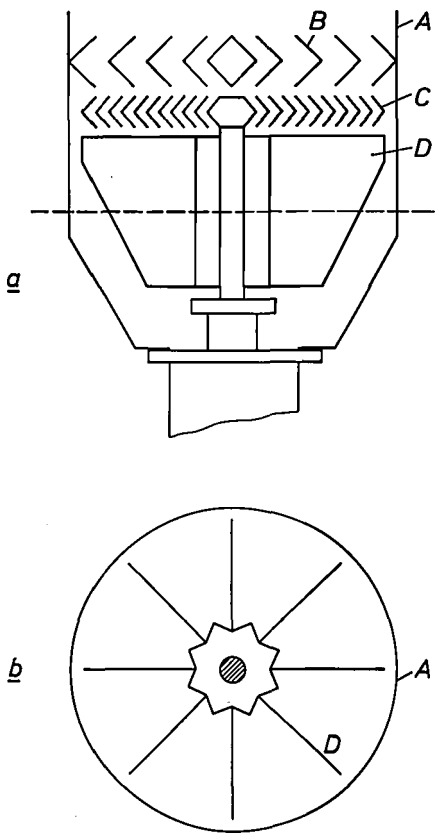


Fig. 4. The configuration of the panels. *a*) Longitudinal section. The baffles A and B are connected to the first stage of the machine and have a temperature of 80 K. The panels C and D are connected to the second stage and have a temperature of 20 K. A is intended to reduce the thermal radiation from the surroundings as much as possible. Baffle B is mainly for pumping H_2O, baffle C mainly for N_2, O_2 and CO_2, while the non-condensable gases (mainly H_2) are adsorbed on the group of radial panels D, which are coated with activated charcoal. *b*) Cross-section through the panels at the dashed line in (*a*).

[2] J. P. Hobson, J. Vac. Sci. Technol. **10**, 73, 1973.
[3] G. J. Haarhuis, Le Vide 29, 351, 1974.
[4] G. Prast, Philips tech. Rev. 26, 1, 1965.
[5] J. Visser, B. Symersky and A. J. M. Geraerts, Vacuum 27, 175, 1977.

Since the charcoal layer has a thickness that is equal to the mean grain diameter, the maximum amount of charcoal that can be applied — with the given panel configuration — is proportional to the grain size. On the other hand a small grain diameter gives a more uniform distribution of the adsorbed gas over the charcoal, and in addition a lower surface temperature is obtained. *Fig. 6* shows the measured relation between the pressure p and the quantity of adsorbed hydrogen q in Pa·l per gram of carbon. The quantity q is not the final value, but the quantity of H_2 that has been adsorbed in the pump after 10 minutes of gas feed at constant pressure. It is found that in 10 minutes at higher pressures charcoal with a grain size of 1 mm takes up a quantity of gas closely approaching the theoretical value. Since the adsorption capacity after 10 minutes for 1 mm grains in the lower pressure range is more than twice that for 2 mm grains, the preferred value for the grain size is 1 mm.

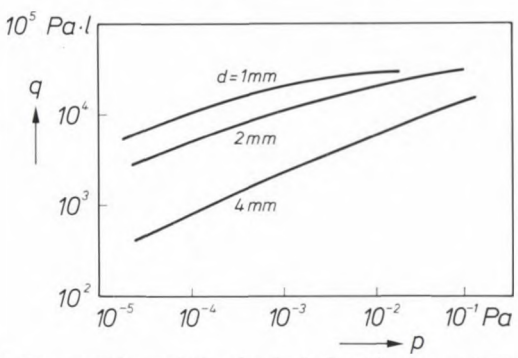


Fig. 6. The quantity q of H_2 adsorbed after 10 minutes, in Pa·l per gram of carbon, as a function of the pressure p, for various diameters d of the grains in the charcoal layer on the radial panels D in fig. 4. As the H_2 is admitted the pressure in the pump is kept constant. The preferred grain size is 1 mm, which gives a better curve in the low-pressure range.

Applications

Pumping a gas-analysis system

We indicated earlier that we would discuss a number of applications of cryopumps, in order of descending pressure. The first application — in which the cryopump is not directly connected to the region at high pressure — is intended for the pressure range of 1 Pa and above. We are concerned with these relatively high pressures in systems intended for processes that take place in a gaseous atmosphere, such as plasma etching, chemical vapour deposition, etc. The knowledge of the gas composition in the process chamber is important for following the process and possibly for controlling it. A mass spectrometer is generally used for this gas analysis. The application that will now be described relates to the pumping of the mass spectrometer connected to the process chamber.

A mass spectrometer consists of an ionization unit, in which the gas molecules to be analysed are ionized, an analysis unit, in which the ionized particles are sorted by mass, and a collector unit, in which the sorted particles are collected. If the mass spectrometer is to operate properly — the ionized particles must not collide with the surrounding gas molecules — the pressure in the analysis unit must not be too high, usually not higher than about 10^{-3} Pa. In many cases it is therefore necessary to reduce the pressure to the desired level in the spectrometer. This is called 'gas sampling'. The gas-reduction systems that have so far been applied have the disadvantages listed below.

— The gas molecules enter into too frequent contact with the walls, etc., so that adsorbed molecules may

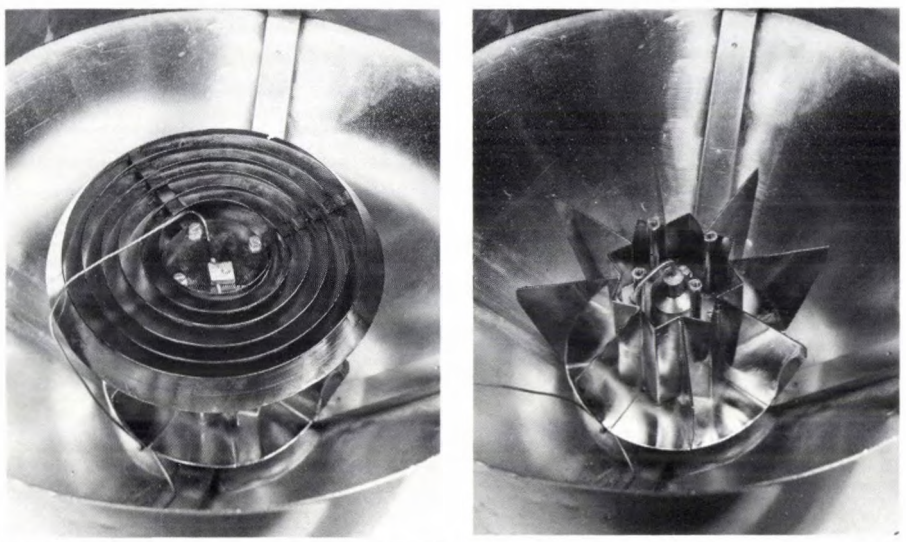


Fig. 5. *Left:* the panels C in fig. 4. *Right:* the radial group of panels D, before coating with activated charcoal.

be detached. Moreover, reactive constituents from the gas — present for example in plasma etching — may react with the walls.

— Too many gas particles are lost, and cannot therefore contribute to the output signal. This is a particular difficulty when appearance-potential mass spectrometry is applied, since a higher sensitivity is then necessary.

In appearance-potential mass spectrometry the gas molecules are not ionized by electrons with the usual constant energy level of 70 eV, but this energy level is reduced in steps, associated with a decreasing production of ions. The threshold value, i.e. the energy level at which ionization just appears, provides information about the molecular structure as well as the molecular mass. This information can be used to distinguish between coincident peaks in the mass spectrum that originate from different ions with the same mass-to-charge ratio.

Fig. 7 shows the system that we have designed, which enables the analysis of gases at high pressures to be carried out more successfully than in conventional systems, even if reactive particles are present. The chamber containing the gas of unknown composition and the spectrometer chamber are connected by a flow aperture *H*. The spectrometer chamber is pumped by a pump consisting of a pair of cryopanels, which are connected to the first and second stages of the cryogenerator and have temperatures of 80 and 20 K respectively. The quadrupole mass spectrometer that we used is almost completely enclosed by these baffles. Since in this type of spectrometer there are no particular requirements for the angle at which the ions enter the analysis unit, the pressure in the ionization unit can be high with respect to that in the analysis unit. The spectrometer was mounted at right angles to the main direction of the gas flow entering the spectrometer chamber via the flow aperture. The gas molecules that miss the ionization unit or do not become ionized land on the cold panels around the spectrometer, so that no reactions with the walls take place and no other particles are detached. Since the pressure is low in and around the analysis unit, it is possible to work with a relatively large gas flow through the ion source, so that the aperture in the wall can be relatively large, about 0.1 mm in diameter. With this high gas flow there is less interference from the background spectrum of the gas atmosphere present in the spectrometer chamber. In addition, the interfering contributions from gases that are desorbed in the ion source ('electron-induced desorption') are proportionately smaller.

We made experiments to find the gas flow at which the system described here still functions well enough to permit the detection of traces of residual gas in the inflow. A flow of 37 Pa·l/s of N_2 contaminated by 100 ppm of hexane was admitted. *Fig. 8* shows part of the mass spectrum, plotted by the recorder attached to the spectrometer. It can be seen that the peaks in the trace, associated with the characteristic breakdown products of C_6H_{14}, stand out so clearly from the background that a concentration ten times as small can be detected. This measurement limit is not

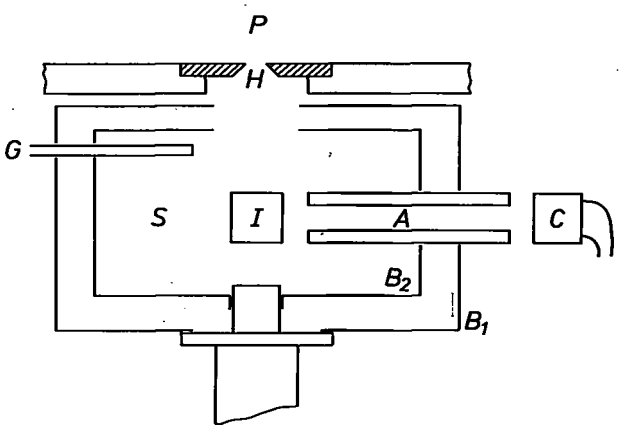


Fig. 7. The pressure-reduction system with mass spectrometer and cryopump. *P* chamber containing gas whose composition is to be analysed. *S* spectrometer chamber. *I* ionization unit, *A* analysis unit, *C* collector unit of the quadrupole mass spectrometer. B_2 baffle connected to the second stage of the cryogenerator, B_1 baffle connected to the first stage. B_2 is at 20 K, B_1 at 80 K. *G* Bayard-Alpert ionization gauge for monitoring the pressure in the spectrometer chamber. *H* flow aperture in the wall to ensure that the pressure in the chamber *P* is reduced to the desired level in the spectrometer chamber *S*.

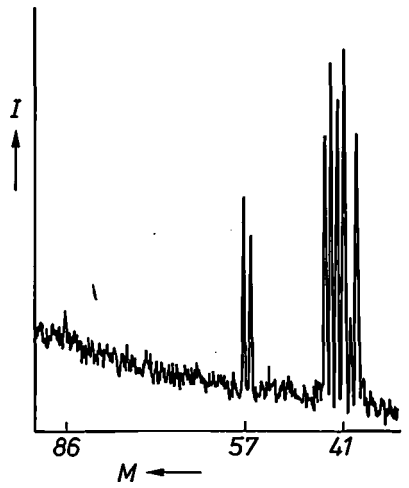


Fig. 8. Part of a mass spectrum, recorded with the system of fig. 7. *I* output signal from the mass spectrometer, as indicated on the vertical scale of the recording device. *M* mass-to-charge ratio of the ion. The part of the spectrum recorded was obtained on admitting 37 Pa·l/s of N_2 to the spectrometer chamber, contaminated by 100 ppm of C_6H_{14}. The peak at $M = 86$ from the hexane itself, of charge 1, can be seen, and also a series of peaks in the neighbourhood of $M = 57$ and $M = 41$; these represent breakdown products of hexane. It is clear that these characteristic peaks could still be distinguished from the background noise even at a tenth of the concentration, so that 10 ppm of hexane is certainly detectable.

imposed by the pressure-reduction system that we designed, but by the (simple) quadrupole mass spectrometer that we employed.

The deposition of thin films by magnetron sputtering

The second kind of application relates to pumping a chamber in which thin films can be deposited on a workpiece by sputtering. The pressures generally used in this process are in the range from 1 to 10^{-2} Pa.

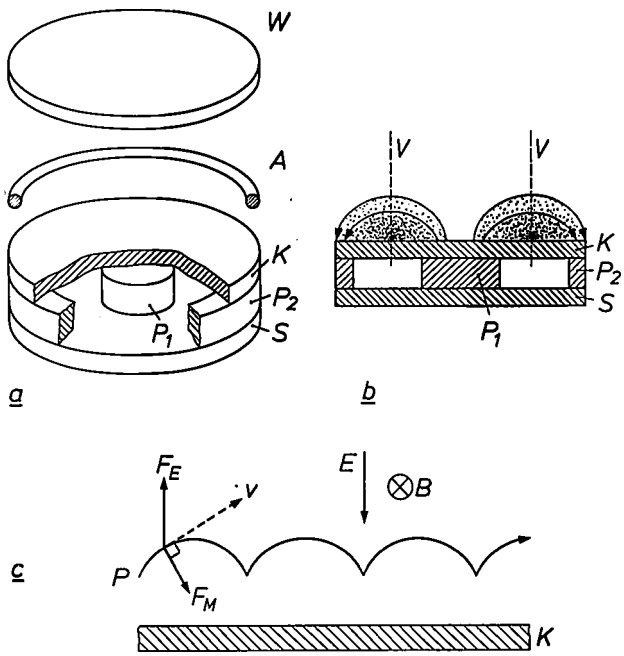


Fig. 9. The principle of magnetron sputtering. *a*) Diagram of the system. The material to be sputtered is placed on the cathode *K*, and a direct voltage is applied between *K* and the anode *A*. When the gas discharge is ignited the field *E* directly above the cathode has the direction shown in (*c*). Below the cathode there are permanent magnets P_1 and P_2, connected by the soft-iron plate *S*. *W* substrate mount. *b*) Central cross-section, showing a few magnetic lines of force. The plasma is concentrated in the hemitoroidal space, shown stippled. *c*) The plane of the diagram is the developed cylindrical surface *V* shown in (*b*). The electric field *E* and the magnetic flux density *B* are perpendicular to one another. An electron moving at velocity *v* is subject to forces F_E ($= e \times E$) and F_M ($= B \times v \times e$). It can be shown that the electron leaves the initial position *P* to describe a series of cycloids. The electron cannot therefore escape from the region of the magnetic field unless there are collisions. The sputtering takes place because the Ar atoms ionized by the electrons strike the cathode and detach particles of the material. These particles strike the workpieces on the substrate mount *W*.

Sputtering depends on the use of a gas discharge in an atmosphere (usually) of argon. The material to be sputtered (the target) is applied to the cathode. The argon ions from the plasma formed in the discharge are drawn to the cathode and collide with the target. This causes particles to be detached from the cathode and these in turn land on the workpiece.

One particular form of sputtering is magnetron sputtering; the principle is shown in *fig. 9*. The plasma concentrates in a hemitoroidal region above the cathode, producing a high density of free electrons. The electrons can only reach the anode after many collisions with gas molecules, so that many ions are produced.

The advantage of sputtering over the deposition of thin films by evaporation is that films consisting of mixtures or alloys can be applied more easily. Another advantage in some applications is that the material to be deposited does not have to be brought to the liquid state. Since in sputtering the particles strike the surface of the workpiece at a higher velocity, the adhesion between the film and the base material is often better.

The problem that arises in pumping the chamber in which the sputtering is carried out is that on the one hand a certain pressure of the argon gas must be maintained (about 0.4 Pa) to maintain the discharge, whereas on the other hand contaminants such as H_2, H_2O and hydrocarbons — which can affect the quality of the deposited film — must be reduced to an acceptable level. In most cases transport pumps such as diffusion or turbomolecular pumps have been used previously. Since these pumps cannot provide stable operation at the required argon pressure, an extra pump resistance or 'throttle' must be included, but this affects the pumping speed for contaminants as well.

Recently cryopumps have been introduced as an alternative to these transport pumps. The types usually used have a fairly limited refrigerating power, so that the quantity of gas that can be removed must still be limited by throttling. This means that contaminants

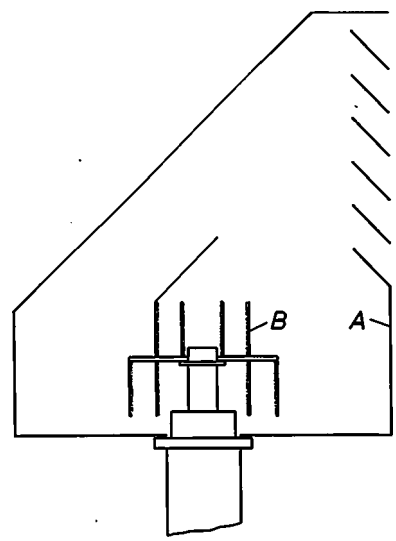


Fig. 10. The cryopump for magnetron sputtering. The panels *A* are connected to the first stage of the K20 cryogenerator and have a temperature of 80 K. These panels pump the readily condensable gases such as H_2O and shield the panels *B* from radiation from the relatively warm surroundings. The panels *B* are connected to the second stage of the cryogenerator and have a temperature of about 20 K. Activated charcoal is applied to the inner side; the charcoal adsorbs the H_2 to be pumped. The panels are again spaced well apart to give the best possible distribution of the pumped gas over the panels.

cannot be removed any more successfully than with transport pumps. We have now constructed a cryopump, based on the K20 cryogenerator mentioned earlier, whose refrigerating power at low temperature is high enough to permit unthrottled operation. The cold panels have to be dimensioned in such a way that the pressure of the argon is not too low, yet nevertheless the pumping speed for hydrogen is adequate.

Since the cryogenerator has a refrigerating power of 10 W at 20 K, it follows that a gas flow of about 2500 Pa·l/s can be pumped. With a pressure of 0.4 Pa for the gas in the sputtering chamber this means that the pumping speed can be 6000 l/s. A pumping speed as high as this is found to be satisfactory for removing contaminants from the sputtering chamber. The pumping-speed equation (4) derived earlier for cryopumps gives a pumping speed for Ar of about 10 l/s per cm^2 of cold surface. A projected surface area of 600 cm^2 is therefore required for the 20 K panels. To obtain a very adequate pumping speed for H_2 and taking account of the resistance of the surrounding baffles (under molecular-flow conditions) we made the panel surface area equal to 1200 cm^2; this also enabled us to apply more sorption material.

In a commercially available system for sputtering films such as aluminium/silicon we replaced the diffusion pump by the cryosystem [6] shown in *fig. 10*. *Fig. 11* shows that the partial pressure of the various components of the gas atmosphere has hardly changed after admitting argon and hydrogen during a number of sputtering runs. In spite of the deposition of large quantities of Ar and H_2 the pumping speed of the cryopump for H_2 and other gases was unchanged.

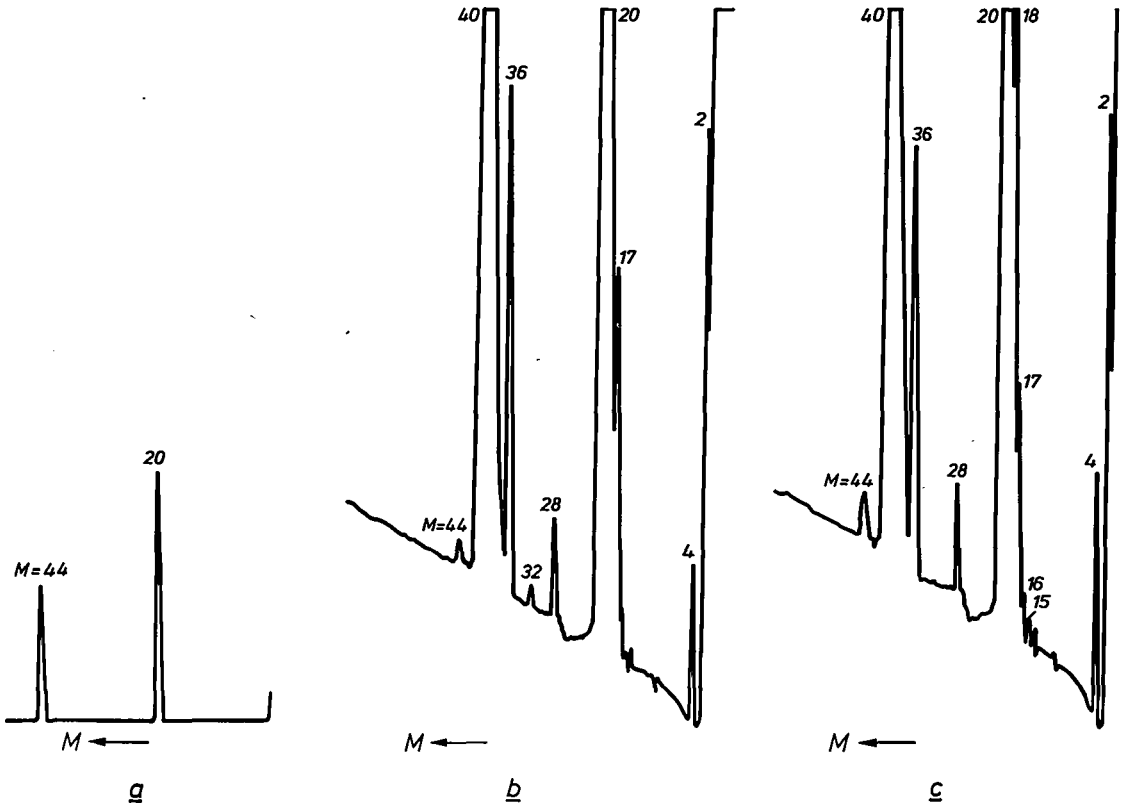


Fig. 11. Some mass spectra of the residual gases in the sputtering chamber. *a*) Part of the spectrum shown with a small expansion of the vertical scale, so that only the peaks for the Ar^+ and Ar^{++} ions ($M = 40$ and 20) are visible. *b*) Part of the mass spectrum of the gas atmosphere, with greater expansion of the vertical scale, immediately before the start of the sputtering process at the beginning of the experiments. *c*) Part of the spectrum, recorded during the sputtering, after a number of sputtering runs (depositing aluminium/silicon) lasting 7 hours. In these runs 1.7 Pa·l/s of H_2 was continuously admitted and a total of 2.3×10^7 Pa·l of Ar and 5×10^4 Pa·l of H_2 was taken up. The expansion of the vertical scale of (*b*) and (*c*) is such that the peaks at $M = 20$ and $M = 40$ are well off the scale. Even though large quantities of H_2 and Ar were taken up by the cryopump, the pumping speed for the various gases remained unchanged after 7 hours. The pressures of CH_4 ($M = 16$ and 15), CO ($M = 28$) and CO_2 ($M = 44$) increased slightly because of desorption during the sputtering; the pressures of H_2O ($M = 18$ and 17) and O_2 ($M = 32$ and 16) decreased slightly since these gases react with Al during the sputtering. The pressure of H_2 ($M = 2$) is found to remain virtually unchanged at about 2×10^{-4} Pa.

[6] J. Visser and J. J. Scheer, J. Vac. Sci. Technol. **16**, 734, 1979.

Depositing thin films by evaporation

In the application considered here the pressure range is 10^{-2} to 10^{-6} Pa. Evaporation is the oldest method of applying thin films in vacuum. In sputtering the coating material remains solid, but in evaporation it is usually brought to the liquid state. The pressure can now be much lower, since no gaseous atmosphere is necessary for the production of charged particles and an adequate mean free path can be obtained. The pressure used in the process is determined by the degree of contamination permissible in the evaporated film.

It is found that considerable quantities of hydrogen are evolved in many evaporation processes. The most important hydrogen-producing processes will now be listed.

— Desorption of contaminants present on the walls. This happens mainly at the start of the process.

— Outgassing of the material to be evaporated. This also happens mainly at the start, for the materials Mo, Cr, Ti, etc., which are often used as getters because of their readiness to combine with gases.

— Dissociation of H_2O at hot surfaces.

— In some cases, a reaction between H_2O and the film being produced.

Table I. The variation in the pumping speed for hydrogen during a series of pumping runs, for different values of the refrigerating power.

Run	Pumping speed for H_2 (l/s)		Refrigerating power (W)	Cumulative quantity of gas taken up ($\times 10^3$ Pa · l)	
	at 4×10^{-4} Pa	at 10^{-4} Pa	at 20 K	H_2	N_2
1-13	4600	4600	9	26	208
14-17	4400	4200	6	33	256
18-29	4200	2500	3	48	432
30-34	3350	0	1	59	496
35-37	4490	4250	9	68	544

Since in practice the evaporation runs are made as short as possible and the system cannot therefore be baked out, the gas in the system consists mainly of H_2O when the evaporation process is started. This means that it is hard to avoid the last two hydrogen-producing processes. The H_2 evolution is greatest when aluminium is being evaporated, mainly because of the reaction between H_2O and the Al film.

We modified a commercially available evaporating system by fitting a cryopump with excellent pumping characteristics for H_2; see fig. 4. The contamination by hydrocarbons from the backing pump was again almost entirely eliminated by starting the cryopump when the pressure had fallen to about 40 Pa.

To investigate the influence of the refrigerating power on the pumping capacity, a number of pumping runs were carried out with different refrigerating powers. Nitrogen contaminated by H_2 was admitted at the start of each run at a pressure of about 40 Pa and the evaporation chamber (volume 300 l) was then pumped down to about 10^{-4} Pa. The refrigerating power was set to values of 9, 6, 3 and 1 W. The results of the various measurements are shown in *Table I*, with the pumping speeds indicated for the last run of each series. It is found that if large quantities of gas are taken up by the cryopanels, the pumping speed at a low refrigerating power may even be reduced to zero. This is because the gas taken up forms a heat-insulating layer, so that the surface temperature increases. If the refrigerating power is restored to its original value of 9 W, the pumping speed returns almost to its initial value. As in magnetron sputtering, the high refrigerating power of the cryogenerator we used was found to be an advantage.

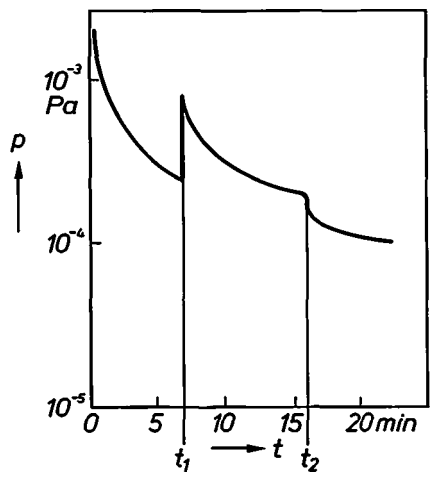


Fig. 12. The pressure as a function of time in the 95th of a series of evaporation runs. At time t_1 the electron gun is switched on, at time t_2 it is switched off. The increased desorption during evaporation can clearly be seen. The material evaporated is aluminium. In between each two successive phases of the evaporation the bell-jar is exposed to atmospheric pressure for 5 minutes while the coated workpieces are removed and new ones introduced. The steep slope of the pumping curve indicates that the pumping speed is still high even though large quantities of gas have been taken up by the cryopump (see fig. 4).

In another series of experiments aluminium was evaporated many times in the system, with the material to be evaporated heated by an electron gun. *Fig. 12* shows the pressure variation during the 95th evaporation run. It can be seen that even though the cryopump has taken up large quantities of gas, the pumping speed is still high.

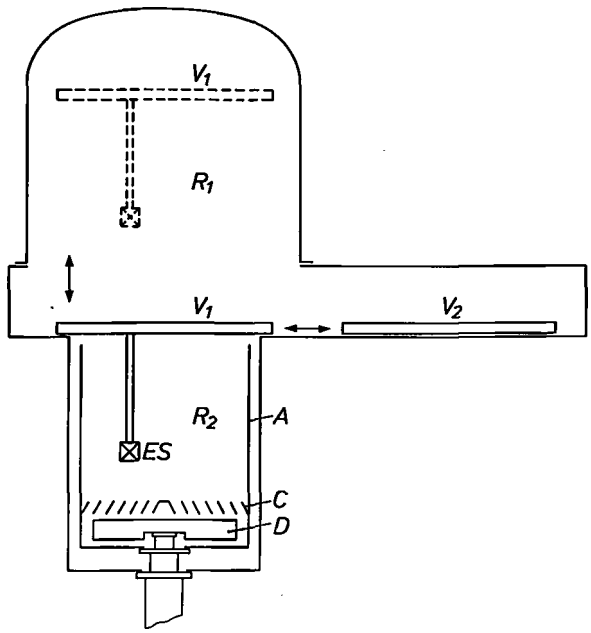


Fig. 13. System for rapid pumping down to ultra-high vacuum. *A*, *C* and *D* are cryopanels connected to the K20 cryogenerator. The panel *A* that surrounds much of the working chamber R_2 and the baffle *C* have a temperature of 80 K, and the panels *D*, which are coated with activated charcoal, have a temperature of 20 K. At the start of a run, e.g. for evaporating on to workpieces with the aid of an electron source *ES*, the chamber R_2 is closed off by the valve V_2. The chamber R_1 is pumped down to a pressure of about 10 Pa by a rotary backing pump. Then V_2 is moved to the right and V_1 downwards, until V_1 in turn shuts off the chamber R_2. Since the workpieces and the electron gun are now almost entirely surrounded by cold walls, a low pressure is soon reached without having to bake out the walls of R_2.

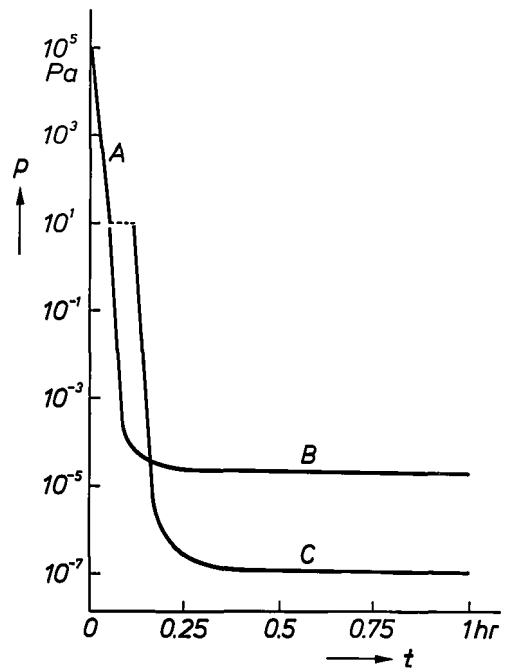


Fig. 14. The pressure in the neighbourhood of the workpieces attached to valve V_1 (see fig. 13) as a function of time. Line *A* shows the pressure variation until valve V_2 is opened: pumping by the rotary backing pump. Line *B* shows how the pressure varies when the evaporation chamber is permanently connected to a cryopump in the conventional way. Line *C* shows that the 10^{-7} Pa range is reached in a short time when the system shown in fig. 13 is used. The short dashed line between the curves *A* and *C* represents the time necessary for the displacement of valves V_1 and V_2.

Accelerated pumping to ultra-high vacuum

The last application that we shall describe is used in the pressure range below 10^{-6} Pa. Previously, if it was desired to reach a pressure in this range within a reasonably acceptable time, it was necessary to bake out the walls of the vacuum system. This is because the desorption of the gas on the walls — which largely determines the gas flow that the pump must handle in the UHV range — is an exceedingly slow process at room temperature: it may take several days. However, baking out takes a long time too, because of the time taken in the heating and cooling of the walls. Moreover, some of the materials in or on the evaporation chamber may be adversely affected by the heating.

A method that we have proposed for reaching low pressures in a reasonable time is illustrated in *fig. 13*. The system shown is intended for evaporating thin films by means of an electron gun, but the principle can also be used for other applications where it is highly desirable to change the substrate quickly, e.g. in equipment for surface analysis. The evaporation source and the workpiece mount, together with the electron gun, can be 'immersed' in a chamber in which it is almost completely surrounded by panels at a temperature of 80 K. In the same chamber there are also cryopanels at a temperature of 20 K that are screened by an 80 K baffle. Since almost all the gas-emitting walls are located behind cryopanels, the flow of residual gases to the workpiece mount is very small. It takes 20 minutes for a pressure of 10^{-7} Pa to be reached in the neighbourhood of the workpiece. This is illustrated by *fig. 14*, which shows the variation from atmospheric pressure (lines *A* and *C*; the short dashed line represents the time necessary for the displacement of valves V_1 and V_2). The second curve (line *B*) in this figure represents the pressure variation in a normal evaporating system connected to a cryopump in the conventional way.

Measurements made with a mass spectrometer have shown that the residual gas in the neighbourhood of the workpiece mount in this system consists mainly of water vapour. This is not the case in baked-out ultra-high-vacuum systems, where the residual gases are mainly H_2, CO and CO_2.

Summary. The ability to attain a hydrocarbon-free vacuum and its high pumping speed make the cryopump of interest for many applications in industrial vacuum technology. A description of the operation of cryopumps is followed by a brief discussion of some of the design aspects. Four applications are then described, each representative of a particular pressure range. These are: pumping a gas-analysis system (> 1 Pa), a system for applying thin films by sputtering (1 to 10^{-2} Pa), deposition of thin films by evaporation (10^{-2} to 10^{-6} Pa) and a method of accelerated pumping down to ultra-high vacuum (< 10^{-6} Pa). The Philips K20 cryogenerator was used in all these applications; its high refrigerating power is a great advantage in some applications and sometimes a necessity.

Scientific publications

These publications are contributed by staff of laboratories and plants which form part of or cooperate with enterprises of the Philips group of companies, particularly by staff of the following research laboratories:

Philips Research Laboratories, Eindhoven, The Netherlands	*E*
Philips Research Laboratories, Redhill, Surrey RH1 5HA, England	*R*
Laboratoires d'Electronique et de Physique Appliquée, 3 avenue Descartes, 94450 Limeil-Brévannes, France	*L*
Philips GmbH Forschungslaboratorium Aachen, Weißhausstraße, 51 Aachen, Germany	*A*
Philips GmbH Forschungslaboratorium Hamburg, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany	*H*
Philips Research Laboratory Brussels, 2 avenue Van Becelaere, 1170 Brussels (Boitsfort), Belgium	*B*
Philips Laboratories, N.A.P.C., 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A.	*N*

C. Belouet & W. T. Stacy: X-ray topographic study of growth sector strains in KDP single crystals.
J. Crystal Growth **44**, 315-319, 1978 (No. 3). *L, E*

S. Boronkay & M. Steers: Memory light valve: smectic-A liquid crystal associated to a photoconductor.
Ann. Physique **3**, 339-344, 1978 (No. 2/3/4). *L*

D. Bouwhuis & H. Bouma (Institute for Perception Research, Eindhoven): Visual word recognition of three-letter words as derived from the recognition of the constituent letters.
Perception & Psychophysics **25**, 12-22, 1979 (No. 1).

A. M. van Diepen & Th. J. A. Popma: Temperature dependence of the hyperfine field in amorphous Fe_2O_3.
Solid State Comm. **27**, 121-125, 1978 (No. 2). *E*

A. G. Dirks & J. R. M. Gijsbers: Crystallisation of amorphous rare-earth–iron and transition-metal–boron thin films.
J. Physics D **12**, 149-154, 1979 (No. 1). *E*

B. J. Fitzpatrick, R. N. Bhargava, S. P. Herko & P. M. Harnack: Near band-edge luminescence in ZnSe grown from gallium solution.
J. Electrochem. Soc. **126**, 341-343, 1979 (No. 2). *N*

D. Hennings & W. Mayr: Thermal decomposition of (BaTi) citrates into barium titanate.
J. solid State Chem. **26**, 329-338, 1978 (No. 4). *A*

G. Jacob, M. Boulou & D. Bois: GaN electroluminescent devices: preparation and studies.
J. Luminescence **17**, 263-282, 1978 (No. 3). *L*

G. D. Khoe & G. Kuyt: Realistic efficiency of coupling light from GaAs laser diodes into parabolic-index optical fibres.
Electronics Letters **14**, 666-668, 1978 (No. 20). *E*

U. Killat & D. R. Terrell: Conventional and self-developing photothermoplastic devices: sensitivity and rapid development.
J. photogr. Sci. **26**, 183-188, 1978 (No. 5). *H*

J. E. Knowles: Measurements on single magnetic particles.
IEEE Trans. **MAG-14**, 858-860, 1978 (No. 5). *R*

J. Lohstroh & N. C. de Troye: Propagation delay time of i.i.l. at low current levels.
Electronics Letters **14**, 705-706, 1978 (No. 22). *E*

A. R. Miedema & F. J. A. den Broeder: On the interfacial energy in solid-liquid and solid-solid metal combinations.
Z. Metallk. **70**, 14-20, 1979 (No. 1). *E*

A. E. Morgan & H. W. Werner: Molecular versus atomic secondary ion emission and its implications for quantitative analysis.
J. Microsc. Spectrosc. électron. **3**, 495-496, 1978 (No. 5). *E*

J. H. Neave & B. A. Joyce: Structure and stoichiometry of {100} GaAs surfaces during molecular beam epitaxy.
J. Crystal Growth **44**, 387-397, 1978 (No. 4). *R*

A. van Oostrom: Review of surface analysis techniques.
Proc. VIII Int. Symp. on Discharges and electrical insulation in vacuum, Albuquerque 1978, pp. C1/1-37. *E*

Ph. Piret: Generalized permutations in convolutional codes.
Information and Control **38**, 213-239, 1978 (No. 2). *B*

I. Sander: A simple optical memory with mechanical access.
Large scale integration (4th Euromicro Symp., Munich 1978), ed. H. W. Lawson, Jr., H. Berndt & G. Hermanson, pp. 154-158; North-Holland, Amsterdam 1979. *H*

M. J. Underhill & P. A. Lewis: Quiet tuning of antennas.
Electronics Letters **15**, 37-38, 1979 (No. 1). *R*

J. P. Woerdman: Self-broadening of the Na 3s-5s and 3s-4d two-photon transitions.
Optics Comm. **28**, 69-72, 1979 (No. 1). *E*

Supplement to Philips Technical Review
Beilage der Philips Technischen Rundschau
Bijlage van Philips Technisch Tijdschrift

PHILIPS

September 1981

Recent United States Patents

Abstracts from patents that describe inventions from the following research laboratories that form part of or cooperate with the Philips group of companies:

Philips Research Laboratories, Eindhoven, The Netherlands	*E*
Philips Research Laboratories, Redhill, Surrey RH1 5HA, England	*R*
Laboratoires d'Electronique et de Physique Appliquée, 3 avenue Descartes, 94450 Limeil-Brévannes, France	*L*
Philips GmbH Forschungslaboratorium Aachen, Weißhausstraße, 51 Aachen, Germany	*A*
Philips GmbH Forschungslaboratorium Hamburg, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany	*H*
Philips Research Laboratory Brussels, 2 avenue Van Becelaere, 1170 Brussels (Boitsfort), Belgium	*B*
Philips Laboratories, N.A.P.C., 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, USA	*N*

4 237 496

Device for coding/decoding data for a medium

E. de Niet *E*

In a coding device, a pseudo-ternary signal is derived from a binary data flow. Two of the three values of the latter signal represent the same data status. The three values of the pseudo-ternary signal determine the transition instant between the two values of the status quantity of the medium used The transition instants are advanced and delayed, respectively, by said two values. The other signal value each time produces a nominal transition instant. A status transition then exists between every two successive bit cells. As a result of said mutual time shifts, a bipolar, phase-modulated signal with inserted pilot tone is generated. This signal, furthermore, does not contain a d.c. component and is suitable for saturation storage in a magnetic medium. The extraction of the data is realized by the formation of an exclusive-OR-function between the signal received and the output signal of a phase-locked loop which is cosynchronized by the signal received.

4 238 277

Method of manufacturing a magnetic device

T. W. Bril *E*
L. Postma

A method of manufacturing a magnetic device for propagating and detecting magnetic domains. At least one thin domain layer of a magnetizable material is provided on a nonmagnetic substrate body and hereon are provided a number of layers of which there are at least a detection layer of magnetoresistive material and a layer of electrically insulating nonmagnetic material. The latter is provided with a propagation layer of magnetizable material in a pattern which satisfies the requirements of propagation of magnetic domains and the requirements of detecting magnetic domains. The non-covered parts of at least the layer of electrically insulating, nonmagnetic material and of the layer of magnetoresistive material are successively removed, each time by means of a method which does not substantially attack the pattern of the material present above the relevant layer.

4 238 527

Method of providing metal bumps on an apertured substrate

M. J. C. Monnier *L*
M. A. Monneraye
C. Foucher

Method of providing metal bumps on an apertured substrate and substrate having metal bumps.

4 238 644

Supply circuit for a subscriber's line circuit

E. C. Dijkmans *E*
K.-D. K. Brockmann

A supply circuit for a telephone subscriber's line circuit that includes a voltage source and a supply impedance connected in series with the line to the voltage source for limiting the line current upon short circuit of the line. The supply circuit includes an a.c. voltage generator, an a.c./d.c. converter, and a circuit for coupling the a.c. voltage generator to the line via the supply impedance and the a.c. converter.

4 238 795

Microwave range measuring system for measuring the distance of an object

B. Schiek *H*
W. Schilz
R. Jacobson

Disclosed is a microwave range measuring system for determining the distance to an object by means of the phase difference between a transmitted frequency-modulated wave and the wave reflected by and received from the object. The frequency of the transmitted microwave signal is varied between two fixed frequencies f_u and f_o and the signal reflected by and received from the object is converted by means of a signal converter into an intermediate frequency signal from whose total phase shift the distance is derived.

4 238 820

Variable wave-form converter

G. J. Naaijer *L*

A d.c./a.c. converter synthesizes an alternating voltage by suitably and at the correct instants selectively combining a plurality of direct voltage sources into a variety of series connections by means of a plurality of semiconductor controlled switches and diodes. Polarity inversion is effected with a bridge circuit comprising switches. The converter load is connected to said bridge circuit. In this way it is possible to directly generate a 220 V, 50 Hz a.c. voltage without inductive elements, from solar panels, using accumulator batteries.

4 238 843

Disc-shaped optically readable record carrier used as a data storage medium

M. G. Carasso E
J. J. Verboom
M. R. de Haan

A disc-shaped record carrier having an information track which is divided into a plurality of sectors per track circumference. Each sector is divided into a data section, in which the data can be recorded, and a synchronizing section. This synchronizing section consists of an optically detectable relief structure and comprises an indicator portion and an address portion. The address portion contains the information about the track number and the sector number. The indicator portion serves to define the beginning of the address portion unambiguously and for this purpose has such a relief structure that the indicator signal produced after cooperation with the radiation beam has a frequency which is clearly distinguishable from signal components resulting from the address portion.

---

4 239 587

Method of manufacturing a thin-film magnetic head with a nickel-iron pattern having inclined edges

G. J. Koel E
L. Postma

In the electrodeposition of nickel-iron alloy layers during the manufacture of a thin-film magnetic head it is ensured that the nickel-iron alloy layer has a top layer having a higher iron content than the iron content of the remainder of the nickel-iron alloy layer. Since this top layer etches more rapidly than the remainder of the layer, the nickel-iron alloy layer itself becomes sloped during etching, this sloping configuration presents great advantages in the subsequent coating with a layer of silicon dioxide.

---

4 239 805

Method of depositing a layer of magnetic bubble domain material on a monocrystalline substrate

H. D. Jonker E

In the LPE growth of magnetic bubble domain garnets from a melt the overall dependence of the saturation magnetization on the growth temperature can be adjusted to make the bubble domain collapse field independent of the growth temperature. In the melts, from which these garnets are grown, two types of non-magnetic ions are substituted for the iron ions, each of which ions, if substituted alone, would result in garnet materials having opposite dependence of the saturation magnetization on the growth temperature.

---

4 241 404

Device for computed tomography

P. Lux H

For determining the absorption distribution in a plane of a body by means of a fan of radiation beams from an approximately point-shaped radiation source and in different positions of the body with respect to the radiation source, a detector element comprising only one amplifier is used for a number of adjacent radiation beams or all radiation beams. For determining the intensity profile in each measuring position, the radiation beams are modulated by time functions which form a complete, linearly independent function system (Walsh functions). Modulation is realized by means of an aperture which is constructed as a rotating disc which comprises transmissive and non-transmissive sectors. At each change of a function value, a discrete signal value is derived from the detector signals, so that the total number of signal values equals the number of functions of the function system. The signal values are multiplied by a back transformation matrix and therefrom the intensity profile is recovered. This profile is processed in known manner in order to obtain an absorption image. This principle is also suitable for a two-dimensional fan of radiation beams for the simultaneous determination of the absorption distribution in a plurality of planes.

4 242 117

Method of producing optical fibers for telecommunications

H. M. J. M. van Ass E

When producing a glass fiber, having a graded refractive index profile, by means of the double crucible method, the initial materials being core and cladding glass compositions having mutually different alkali ions, a profile is usually obtained which greatly deviates from the desired parabolic form. The invention furnishes the possibility of approximating this parabolic form very closely. This is attained by a partial substitution of the core alkali ion by the cladding alkali ion.

---

4 242 315

Hydrides of the formula AB_nH_m

H. A. C. M. Bruning E
J. H. N. van Vucht
F. F. Westendorp
H. Zijlstra

A hydride of the gross formula AB_nH_m, in which A is calcium or one or more of the rare earth metals with or without thorium, zirconium, or hafnium, B is nickel and/or cobalt, with or without iron and/or copper, n has a value between about 3 and 8.5 and m has a value up to about 7.

---

4 242 579

Apparatus for the pointwise scanning of an information surface

G. Bouwhuis E

An apparatus is described for the pointwise scanning of an information surface. By arranging an observation objective system asymmetrically relative to a zero-order subbeam of the radiation coming from the information surface, and by utilizing a detector whose dimension in the scanning direction is small, the resolution can be increased.

---

4 242 596

Integrated injection logic circuits

K. R. Whight R

An integrated injection logic circuit having at least one threshold logic gate having a plurality of weighted logic inputs and a weighted fixed input determining the threshold, said gate having first and second transistors connected as a bistable with the collector zone of the first transistor connected to the base zone of the second transistor connected to the base zone of the first transistor, at least one of said first and second transistors having a further collector zone provided with a conductive output connection, a plurality of current injector structures being present for determining the bias currents supplied to the first and second transistors, the state of bistable being determined in accordance with the difference in magnitude of total bias current supplied to the first transistor and total bias current supplied to the second transistor, said plurality of current injector structures being associated with the logic inputs and at least one injector structure associated with a fixed input.

---

4 242 601

Circuit arrangement for frequency division

W. G. Kasperkovitz E

A wide-band frequency divider having a first and a second group of n transistors each, with their emitters in a first cyclic sequence alternately connected to the collectors of two transistors forming an input differential pair. The base electrode of any arbitrary transistor of said first cyclic sequence is always connected to the collector of the $(n-1)^{th}$ transistor after said arbitrary transistor in said first sequence. The frequency divider further comprises a third and a fourth group of n transistors each, also with their emitters alternately connected in a second cyclic sequence to the collectors of two transistors which constitute an input differential pair. The collectors of the transistors in the second sequence are each time con-

nected to the collector of the next transistor in the first sequence. The base electrode of any arbitrary transistor of said second cyclic sequence is always connected to the collector of the n^{th} transistor after said arbitrary transistor in that sequence.

4 242 688

Ink jet printer

M. Döring H

An ink jet printer comprising one or more nozzles which are connected, via a pressure chamber, to an ink supply duct. The pressure chamber comprises a wall which can be moved by means of an electromechanical converter in order to realize droplet-like ejection of ink from the nozzle. The converter continuously vibrates during operation. In front of each nozzle there is arranged a brake electrode which can be connected to a voltage as desired. When the voltage is switched off, the ejected ink droplet is incident on the record carrier, and when the voltage is switched on, the ink droplet is intercepted by the brake electrode and is withdrawn into the nozzle by the converter. Special interceptors for the ejected ink and deflection devices are no longer required.

4 243 399

Method of producing a coupling element for an optical transmission fiber

G. D. Khoe E
R. G. Gossink
C. M. G. Jochem

Method of producing a coupling element for coupling a laser radiation source to a monomode optical transmission fiber. The end of the fiber is monotonically flattened at a temperature at which the fiber's core has a viscosity of between 10^7 to $10^{8.5}$ poises. The fiber has a cladding glass chosen to have a viscosity of between 10^{10}-10^{11} poises at the flattening temperature. This causes the core glass to emerge in the form of a semi-ellipsoidal lens, when the fiber is flattened.

4 243 906

High pressure mercury vapor discharge lamp

I. L. Wilson A

The advantageous light-technical properties of the alkaline earth and rare earth halides can be utilized without an excessive thermal load of the wall of the vessel occurring, if the discharge vessel contains at least one halide of iron, of cobalt or of nickel as a compound-forming metal halide.

4 245 233

Photosensitive device arrangement using a drift field charge transfer mechanism

J. Lohstroh E

A photosensitive element and a photosensitive device arrangement using the element include a charge transfer structure having an electrode layer extending over a photosensitive area of a semiconductor body. In operation, a bias potential is applied to the electrode layer to form a depletion layer in the underlying body portion, and a drift field is produced in the depletion layer which extends in the direction of an edge portion of the electrode layer to permit photogenerated charge carriers to be transmitted towards the edge portion. A preferred structure for producing the desired drift field includes a resistive electrode having first and second connections for applying a potential difference along the resistive electrode. The photosensitive device arrangement further includes a localized charge-storage zone adjacent the edge portion of the electrode layer for collecting the photogenerated charge carriers and a detector circuit for measuring the charge state of the charge-storage zone.

4 246 439

Acoustic writing combination comprising a stylus with an associated writing tablet

J. J. Romein E

An acoustic writing combination includes a stylus with an associated writing tablet. The stylus is provided with two ultrasonic sound sources which upon contact of the stylus with the writing tablet emit pulse-shaped sound signals which are picked up by at least two microphones which are located at the edge of the writing tablet. The two sound sources are situated at different distances from the stylus tip and are operated to alternately produce ultrasonic signals. Point-shaped or circular sound sources may be employed. The circular sound sources may then comprise piezoelectric ceramic rings.

4 246 483

X-ray apparatus for tomosynthesis

H. Weiss H
R. Linde
U. Tiemens
E. Klotz

Apparatus for examining objects includes a group of X-ray sources, which are activated group-wise by a generator. A group of subimages are separately projected onto a photographic film. During a subsequent step, the film is re-imaged with the aid of an optical lens matrix. The lenses in the matrix are arranged in a manner similar to the X-ray sources.

4 248 633

Universal copper-plating solution

G. H. C. Heijnen E
A. Molenaar

A copper-plating bath suitable for reprographic uses and for making electrically conducting metal patterns with one of the redox pairs V^{2+}/V^{3+}, Ti^{2+}/Ti^{3+}, or Cr^{2+}/Cr^{3+} as a reducing agent, ascorbic acid with an acid acceptor, or the redox pair Fe^{2+}/Fe^{3+} together with an organic carboxylic acid and a complexing agent for cuprous ions.

4 249 800

Display device

A. M. J. M. Spruijt E

In a display device having a liquid display medium, for example liquid crystal, the support plates must be kept accurately spaced apart. This is done by means of a monolayer of grains having a hard core and a thermo-plastic envelope.

Re. 30 513

Color splitting prism assembly

A. A. J. Bluekens E
J. Stoffels
P. J. M. Peters

A beam splitting prism assembly in which at least some of the prisms are separated by an air gap, optionally connected at the edges through the beam splitting layers by means of strips of a thermo-plastic material.

4 251 160

Method and arrangement for aligning a mask pattern relative to a semiconductor substrate

G. Bouwhuis E
T. F. Lamboo

A method and arrangement for aligning a mask comprising a mask-pattern relative to a substrate when the mask-pattern is repeatedly and directly imaged on the substrate, gratings in the mask and gratings on the substrate (phase gratings) being employed as alignment references. The gratings in the mask are located outside the mask pattern and the phase gratings are located on the substrate outside the area where the mask-pattern is imaged. The substrate (phase) gratings are imaged on one of the mask gratings with a projection system which is also used for projecting the mask-pattern on the substrate. The image of the gratings on the grating in the mask is adjusted. Thus, a very accurate alignment can be achieved.

4 251 254

Tools for the handling and shaping of glass

J. T. Klomp E
P. J. Vrugt

Tools for the handling and shaping of glass, such as dies, moulds, shears. Sticking of the glass to the tool is prevented by the presence, at the tool surface, of a layer which has a low intrinsic strength with respect to the glass. This layer may be present as such or may be produced after oxidation. A number of possible materials are: boron carbide, Ni-Cr-B-alloys, tantalum boride.

---

4 251 748

Camera tube having photocathode absorptive of shorter wavelength and filter absorptive of longer wavelength light

H. F. J. J. van Tongeren E
P. P. M. Schampers
W. P. Weijland
M. G. Carasso

The optical cross-talk in a camera tube is reduced by the provision of a filter element which overlaps the entire target. The filter may be arranged in front of the photosensitive target and have a spectral transmission which is adapted to the spectral sensitivity of the target. The filter may also be arranged behind the target and preferably constructed so that it is also low-reflective for incoming light. In order to prevent excessive transverse conduction, such a filter may have a mosaic structure consisting of mutually insulated areas. Optical cross-talk may also be reduced by mounting the target on a support which is separately arranged in the camera tube and which has a small thickness.

---

4 251 801

Mobile data communication system

W. Le Mair E
A. N. C. Vereijken

A mobile data communication system in which a fixed station consecutively interrogates a fleet of vehicles to receive and send messages thereto. A receiver of the fixed station effects for each vehicle in turn synchronism with the reply message transmitted by the relevant vehicle. To do this the bit phase error between the reply message signal and a clock signal produced by the fixed station is measured in the latter by means of a phase discriminator and corrected in a control loop comprising also an integrator, an adder and a divider. A store stores information relating to the position of the vehicles, and control means use this information to produce a control signal corresponding to the bit phase error caused by the time delay of the message from the fixed station to the vehicle and vice versa. A switch is operated at the beginning of each message sequence to feed the control signal into the control loop to practically correct the bit phase error.

---

4 251 886

Transmission system for the transmission of data pulses

E. Roza E

A transmission system for the transmission of data pulses from a transmitter to a receiver through a transmission medium, the transmitter comprising a pulse source connected to the transmission medium for applying n-valent pulses to the transmission medium, which pulses, when passing through the transmission medium give rise to the occurrence of intersymbol interference, these pulses furthermore being affected by noise, wherein the receiver includes a pulse repeater, connected to the transmission medium, with quantized feedback.

4 251 909

Method of manufacturing a target assembly for a camera tube

A. M. E. Hoeberechts E

A target and target assembly for a camera tube in which a semiconductor plate is provided on an annular support. The plate has a semiconductor monocrystalline edge portion which comprises an integrated circuit for processing the electrical signals originating from the target. The central portion of the plate is provided with a radiation-sensitive layer having one or more radiation-permeable electrodes. The integrated circuit is provided with inputs which are connected to the electrodes and with leads for the supply and control voltages. A window is provided on the electrodes and overlaps the inner edge of the support, the window, the edge portion and the support adjoining each other in a vacuum-tight manner.

---

4 252 995

Radio broadcasting system with transmitter identification

J. M. Schmidt E
J. O. Voorman

FM radio broadcasting system with transmitter identification by means of a transmitted subcarrier whose frequency is a harmonic of a subharmonic of the stereo pilot and which is binary phase modulated with transmitter identification signals. In the transmitter the subcarrier is locked to the stereo pilot and the unmodulated wave, which is required for detection of the transmitter identification signal, is derived in the receiver from the received stereo pilot.

---

4 253 019

Apparatus for reading an optical record carrier having a radiation-reflecting information structure

W. G. Opheij E
J. E. van der Werf

An apparatus is described for reading a record carrier with an optical radiation-reflecting information structure. The path of the read beam includes a plane-parallel plate, whose surface area is substantially smaller than the cross-sectional area of the read beam, which plate ensures that the beam portion which passes through the plate cannot interfere with the rest of the read beam. After reflection by the record carrier the said beam portion passes through the radiation-deflecting element which deflects the beam portion to two radiation-sensitive detectors. The position of the radiation spot formed by the said beam portion is proportional to the degree of focusing of the read beam on the information structure.

---

4 254 359

Camera tube with graduated concentration of tellurium in target

P. J. A. M. Derks E
J. H. J. van Dommelen
J. Dieleman

A camera tube having an electron source and a target to be scanned on one side by an electron beam emanating from the source. The target comprises a signal electrode and a selenium-containing vitreous layer containing tellurium, in a concentration which varies in the direction of the thickness of the selenium-containing layer, and arsenic. The selenium-containing layer is on the side of the target to be scanned by the electron beam, the signal electrode is on the side of the target to receive radiation. The tellurium concentration in the selenium-containing layer increases from the radiation-receiving side of the layer to the side which is scanned by the electron beam in such manner that over a distance of at most 0.3 microns from the radiation-receiving side, the concentration reaches a value of at least $4\frac{1}{2}$ atomic % and the arsenic concentration uniformly exceeds $1\frac{1}{2}$ atomic %.

PHILIPS TECHNICAL REVIEW

VOLUME 39, 1980, No. 10

Low-noise 12 GHz front-end designs for direct satellite television reception

P. Harrop, P. Lesartre and T. H. A. M. Vlek

Headlines such as 'Experiment with satellite television during Firato' and 'Firato picks up TV from satellite' proclaimed the introduction of a new technology — the reception of television programmes from a distant satellite above the equator — to the Dutch public. The occasion was the 'Firato' consumer-electronics exhibition held in Amsterdam in August 1980; the satellite was the OTS-2 European Orbital Test Satellite, which picked up a number of television programmes transmitted up to it and reradiated them in a broad beam to the whole of Western Europe. The transmissions were received by a three-metre dish antenna on the roof of the exhibition building, and were supplied to a cable-television system.

Future broadcasting satellites will have a higher transmitter power, and with an antenna diameter of a metre individual viewers will be able to receive the programmes. However, the satellite is at a distance of 36 000 km and the signal is attenuated by more than 200 dB during its long journey through space. This means that the receiver front-end has to meet the most exacting requirements. Two laboratory designs are discussed below; these were made so that practical experience could be gained from the reception of the experimental broadcasts from the Japanese television satellite 'Yuri'. These transmissions have now been discontinued and the information collected is being put to use in a pilot design intended primarily for the European market.

Television broadcasting with a satellite

For several years there have been a not insignificant number of satellites circling around the Earth in the service of world-wide telecommunication links. These communication satellites function as relay stations for telephone and television links operating over long distances. They describe an orbit around the Earth once every 24 hours, at a height of 36 000 km above the equator, so that they are always directly above the same point of the Earth's surface, and are therefore said to be 'geostationary'. This is also true for another class of satellites, still in an experimental phase — the broadcasting satellites (*fig. 1*). These are not intended for links between a small number of ground stations, but for transmission of programmes beamed up to them from the Earth to large numbers of individual viewers, who may be using central antenna systems.

P. Harrop, Ph.D., and P. Lesartre are with Laboratoires d'Electronique et de Physique Appliquée (LEP), Limeil-Brévannes (Val-de-Marne), France; Ir T. H. A. M. Vlek is with the Philips Video Division, Eindhoven.

The arguments for satellite broadcasting are of two kinds:

— More programmes can be transmitted for each country than the frequency spectrum for terrestrial broadcast television will permit. Multiple use of each satellite channel is possible as each country will point its receiving antennas at its own satellite, which will direct its signals mainly towards its own territory.

— A satellite enables 100% of the intended service area to be reached at a stroke.

The first argument is particularly applicable for densely populated Western Europe with its many language areas. The second is true for mountainous countries and for those with extensive territory such as Canada, which in fact already had an experimental broadcasting satellite in operation, or India which had an American television satellite 'on loan' in 1975/76; the intention was to use the satellite for disseminating programmes that would encourage the use of birth control.

So far no broadcasting satellites have been launched to serve Western Europe, but a basic system of international agreements has been set up. In 1971 a World Administrative Radio Conference held at Geneva appointed three frequency bands for satellite broadcasting: 620-790 MHz, 2.5-2.69 GHz, and 11.7-12.5 GHz. The lowest of these bands lies inside the frequency band for terrestrial UHF television. Progress made in microwave technology in the years that followed enabled broadcasting engineers to direct their attention almost exclusively to the 12 GHz band, where smaller antennas suffice to create narrow beams of electromagnetic waves.

Still higher frequencies are being considered. Use is to be made of the 'windows' here that are left between the absorption lines of H_2O (22 GHz), O_2 (about 60 GHz), and higher absorption lines. Even inside these windows, however, atmospheric absorption is higher than at 12 GHz.

In 1977 at another World Administrative Radio Conference (WARC '77) the frequency band from 11.7 GHz to 12.5 GHz was divided into 40 channels, and five channels were allocated to each West European country, as well as the longitude for a geostationary satellite and the method of polarization for the radiated carrier (*Table 1*). The five channels allocated to a particular country are either in the band

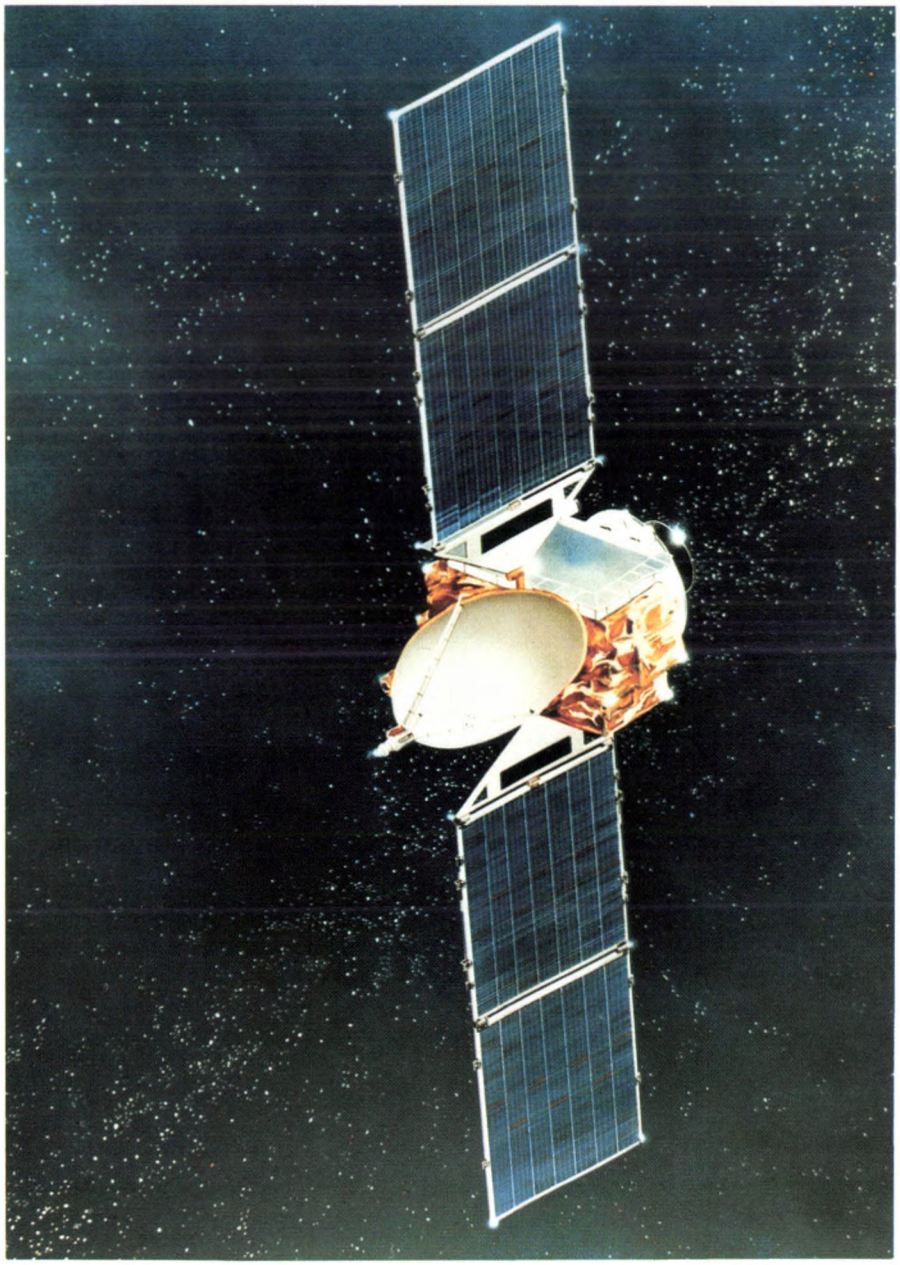


Fig. 1. Artist's impression of the Japanese experimental broadcasting satellite 'Yuri', in operation until eighteen months ago. The transmissions were used for tests in Japan on Philips receivers for satellite television. (Reproduced by permission of NASDA, Tokyo.)

11.7 to 12.1 GHz or in the band 12.1 to 12.5 GHz. Neighbouring countries have opposite circular polarizations. It is possible to provide every television programme with two or more simultaneous sound channels, for example for transmission in more than one language or with stereophonic sound.

Table I. Agreements relating to satellite broadcasting in Western Europe (World Administrative Radio Conference, Geneva 1977). The transmitted signals are to be circularly polarized, 1: right-hand polarization, 2: left-hand polarization. Each country has either five odd channels or five even channels allocated to it. All 40 of the channels are between 11.7 GHz and 12.5 GHz (see also fig. 4).

Country	Longitude of satellite (° EL)	Polarization	Channels				
France	−19	1	1	5	9	13	17
Poland	− 1	2	1	5	9	13	17
Federal Republic of Germany	−19	2	2	6	10	14	18
Finland	+ 5	2	2	6	10	22	26
Luxembourg	−19	1	3	7	11	15	19
Czechoslovakia	− 1	2	3	7	11	15	19
Austria	−19	2	4	8	12	16	20
Great Britain	−31	1	4	8	12	16	20
Sweden	+ 5	2	4	8	30	34	40
Norway	+ 5	2	14	18	28	32	38
Belgium	−19	1	21	25	29	33	37
German Democratic Republic	− 1	2	21	25	29	33	37
Yugoslavia	− 7	1	21	25	29	33	37
Monaco	−37	1	21	25	29	33	37
Switzerland	−19	2	22	26	30	34	38
Hungary	− 1	1	22	26	30	34	38
The Netherlands	−19	1	23	27	31	35	39
Italy	−19	2	24	28	32	36	40
Denmark	+ 5	2	24	28	32	36	40

Each national broadcasting satellite will make its own 'footprint' on the Earth's surface (*fig. 2*). These 'footprints' will also overlap the territory of neighbouring countries; it is estimated that as a result of this 'spillover' many of the French population will be able to receive German transmissions and vice versa, particularly where community receivers equipped with larger antennas are used. The size of the 'footprints' as represented is in fact related to the signal level required for acceptable individual reception.

France and the Federal Republic of Germany are now cooperating in the development of two experimental television satellites. These should be working in pre-operational transmissions in 1983. They will make use of three channels, two for a television programme, the third possibly for high-quality stereo radio. Luxemburg has reserved launching capacity for 1985 and by 1990 a shared satellite ('Nordsat') for the four Scandinavian countries should also be in orbit. The launch will be made with the European carrier rocket Ariane, which was first successfully launched in December 1979 from the Kourou launching base in French Guiana.

To receive satellite broadcasts the viewer must have a 12 GHz receiver. This means that microwave technology, which until now has been almost entirely used in professional applications, has to contend with the constraints set by cheap mass production. We shall give an account of two suitable designs. In quantity production, these receivers would be sufficiently inexpensive for the extra costs (including installation and accurate pointing of the antenna) to be appropriate for the consumer market. But these cost aspects do mean that the satellite will have to have a high transmitter power, up to a few hundred watts per channel. Since this power is transmitted in a very narrow beam (the aperture angle is less than two degrees) it is in fact possible to produce a sufficient field-strength at the receiver antenna even at the enormous distance of more than 36 000 km (*fig. 3*).

The construction of a satellite with such a high transmitter power is no minor piece of engineering. The microwave power is provided by travelling-wave tubes. These derive their d.c. supply from solar cells, which are mounted on large panels whose weight, together with that of the deployment mechanism, contributes considerably to the weight of the satellite. The total weight of the satellite can amount to nearly one tonne. To keep within Ariane's permitted payload the first French and German broadcasting satellites to be produced will only have three channels instead of the permitted five. The transmitter power will be 250 to 330 W per channel. The efficiency of the travelling-wave tubes is 25% to 50%, and for the entire satellite with three channels a d.c. power of 4.2 kW is required. If we assume that the density of the incident solar radiation is 1.35 kW/m^2, the efficiency of the solar cells is 10% and the active proportion of the area is about 75%, then the solar panels will have to have an area of no less than 42 m^2.

The Canadian experimental satellite mentioned earlier — it was called 'Hermes' — had a transmitter power of 235 W; the receivers that we shall describe here were tested with this satellite in Canada (*fig. 4*). Until eighteen months ago the Japanese experimental broadcasting satellite — 'Yuri' shown in fig. 1 was still in operation; it had a transmitter power of 100 W. Four Philips receivers were under test in Japan.

Fig. 2. The 'footprints' that the national broadcasting satellites will make on the territories of the countries of Western Europe. The shape of the footprints is only provisional. It is nevertheless clear that they will overlap the territory of neighbouring countries, thus offering a wider choice of programmes.

The receiver front-end

Its functions

The function of the receiver front-end is to pick up the 12 GHz signal from space with as little noise and interference as possible and to convert it into a signal at a lower frequency which can be subjected to a conversion from the frequency modulation of the satellite signal into the standard television signal. The signal must also be brought to a high enough level, of course.

Several different channels are received from the satellite at the same time. If it is intended to receive a

single channel at a time, as may be the case in private installations, the channel selection is made before the FM demodulator in the receiver. If the receiver is located at the input to a central antenna system, channel selection must be made first, with FM demodulation and AM remodulation for each individual channel later. However, now that private users often have more than one receiver in the home it is sometimes desirable to have parallel operation there too.

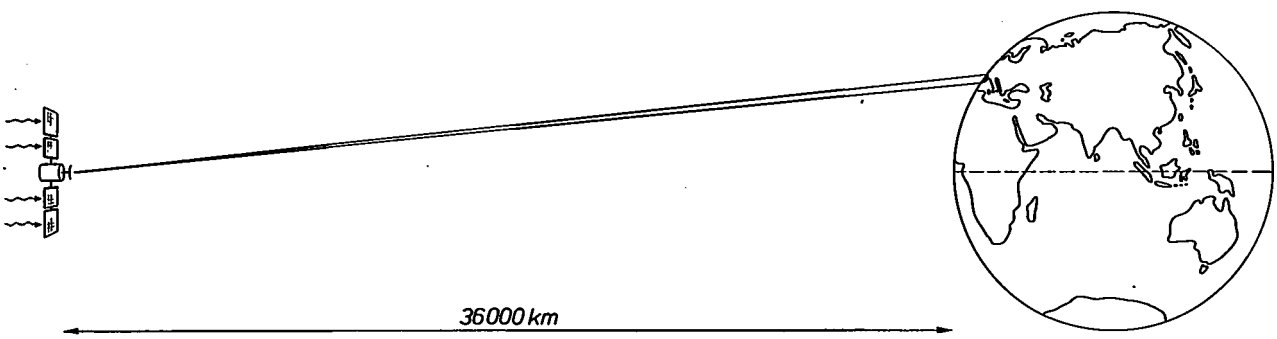


Fig. 3. Scale drawing showing the distance between a geostationary broadcasting satellite and the Earth (the satellite dimensions are not to scale, of course). The beam shown transmits to the whole of France.

The first two countries in Western Europe to have a television satellite will be France and the Federal Republic of Germany. The fact that the spillover enables many viewers to watch programmes from neighbouring countries may be one of the attractions of satellite television. The receivers therefore have to be compatible with this. A glance at Table I shows that France will transmit on odd channels and West Germany on even channels. The channels overlap (*fig. 5*); this is permissible since successive channels are given left-hand and right-hand polarizations alternately. WARC '77 specifies a discrimination of 31 dB between the two directions of polarization at the receiver. The receiver will therefore have to have a system to receive both senses of polarization.

Arrangement of a possible receiver design

The signal from the satellite is received by means of a directional antenna consisting of a parabolic reflector with a diameter of about a metre. At the focus of the reflector there is a horn-shaped input to a waveguide. The depolarizer, which goes inside the waveguide, is still the subject of an investigation.

The waveguide forms an integral part of the housing containing the input stages of the receiver (section *OU* in *fig. 6*). These input stages, which are mounted in the antenna out of doors, include an input bandpass filter, a 12 GHz preamplifier (in one of our designs), a mixer with an 11 GHz local oscillator and an intermediate-frequency amplifier to compensate for the losses in the coaxial input cable to the following sections of the receiver.

The signals in the intermediate-frequency band around 1 GHz are applied to the second section *IU* of the receiver (see fig. 6) via this cable. This section is mounted indoors close to, and later probably in, the television set. In the second section there is first another frequency conversion, to an intermediate frequency of about 120 MHz. At this second intermediate frequency the bandwidth is limited to the 27 MHz required for a single television channel (in FM). Channel selection is made by selecting different oscillator frequencies, which are separated from one another by a spacing of 19.18 MHz, which is possible with an oscillator frequency at about 880 MHz. In a single-channel receiver a single oscillator frequency is selected by means of a selector switch (fig. 6), and in a multichannel receiver there are as many mixers as channels, each with its own local oscillator.

At the second intermediate frequency, amplitude limiting is followed by FM demodulation with the aid of a synchronous 120 MHz demodulator. Next, the demodulated signal is used to modulate the amplitude of a carrier that will be applied to a standard television receiver; the lower side-band is partly filtered out to produce the conventional vestigial-sideband modulation. The 120 MHz oscillator for the FM demodulation is included in a phase-locked loop to enable it to track the frequency of the incoming signal.

Alternative solutions, of course, do exist; they include direct demodulation and threshold extension. These are being investigated.

Two front-end designs

Two designs for the outdoor microwave unit have been produced in the laboratory. The older design (front-end receiver I) has a Schottky-barrier diode

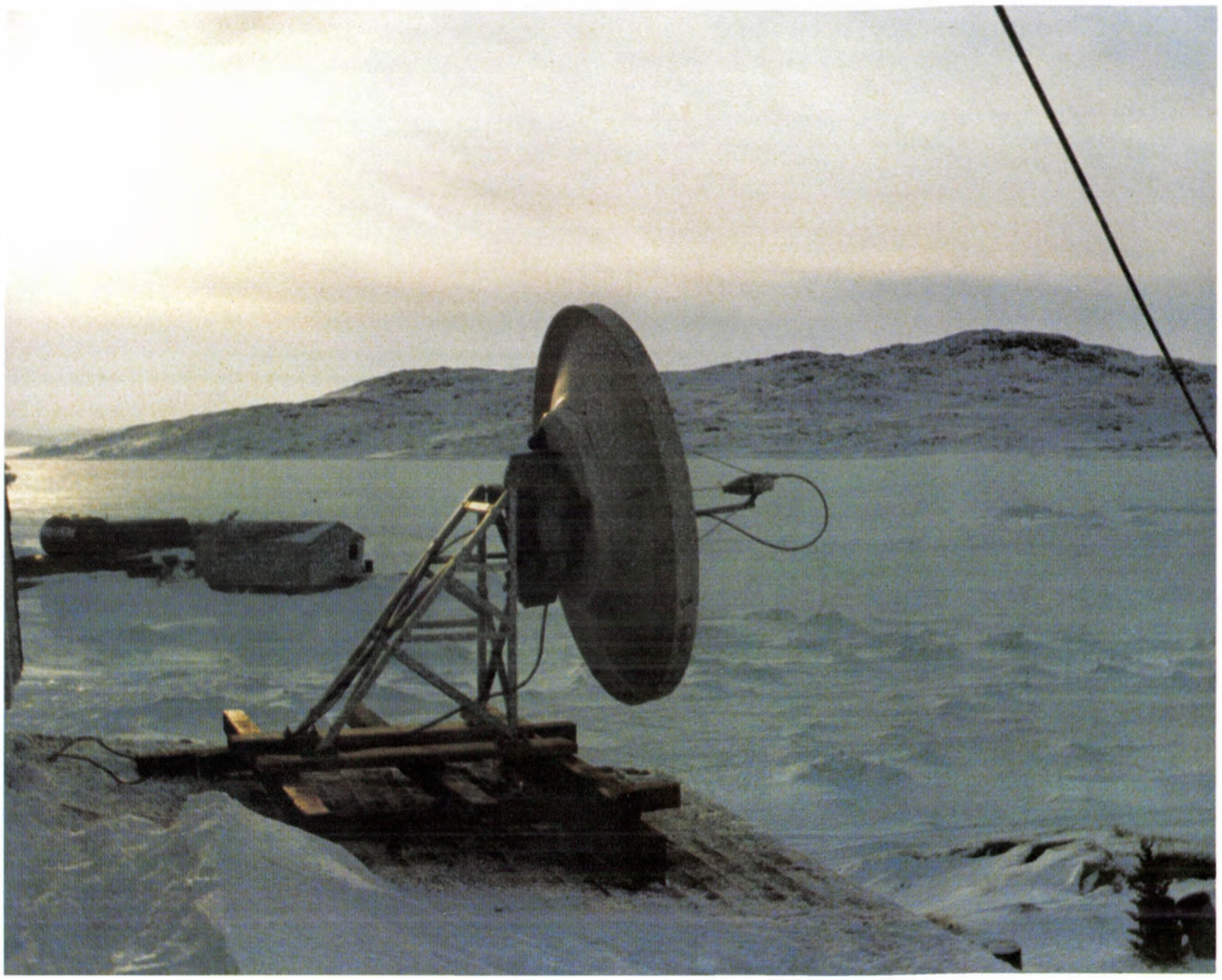


Fig. 4. Experimental arrangement with a 12 GHz receiver in Canada. Television programmes were transmitted by the Canadian experimental satellite 'Hermes'. The picture shows how low the satellite is above the horizon in northern Canada.

immediately following the input bandpass filter. The second design (front-end receiver II) makes use of recently available low-noise field-effect transistors, made in gallium arsenide [1]; this front-end has a two-stage preamplifier before the mixer.

Both designs make use of the same housing incorporating a waveguide for signal input, an arrangement that is suitable for mass production. This housing is made from metallized plastic.

Front-end II has the lower noise figure of the two designs. This is an important aspect in producing a receiver design for satellite television.

Noise, 'link budget'

During its long journey from the satellite to the receiver the transmitted signal undergoes an enormous attenuation: about 10^{20} times (about 200 dB). This is why the signal-to-noise ratio is such an all-important factor in the design of the transmitter-receiver system.

A general picture of the values of the signal and noise power available at various stages is given by a 'link budget'. *Table II* shows such a link budget for the combination of the Japanese experimental satellite 'Yuri' with our two receivers. All power levels are expressed in dB with respect to 1 W.

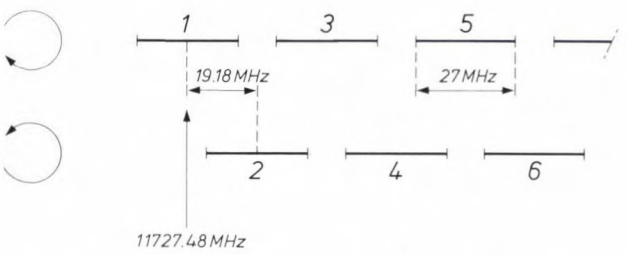


Fig. 5. The channels in the 12 GHz satellite-television band overlap; the spacing is 19.18 MHz and the bandwidth of the channels is 27 MHz. Sufficient channel separation can be obtained by giving successive channels left-hand and right-hand polarization alternately. A particular country has either odd or even channels allocated to it.

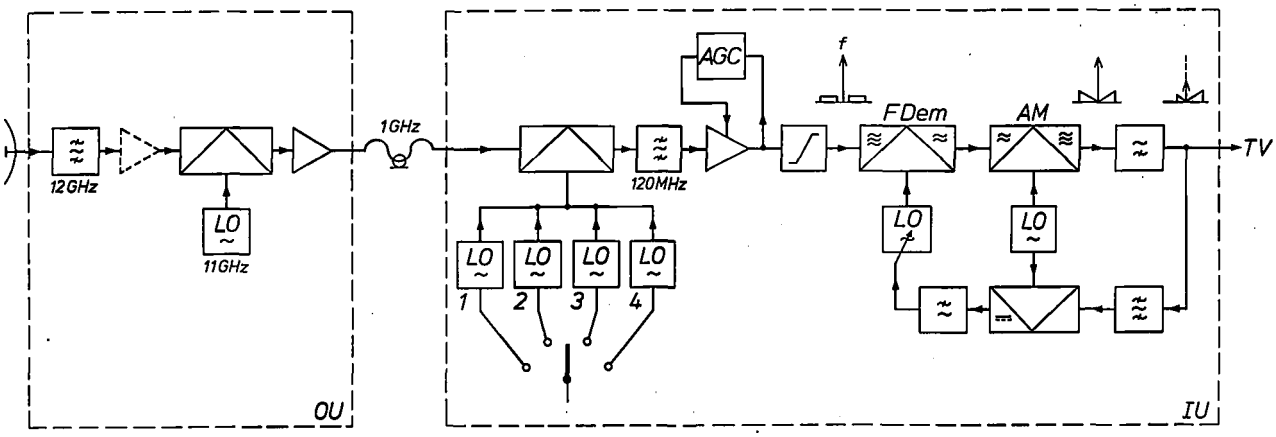


Fig. 6. Block diagram of a 12 GHz receiver for satellite television. The outdoor unit *OU* is mounted directly on the antenna and gives a frequency conversion to a first intermediate frequency of 1 GHz; the converted signal is taken indoors on a coaxial cable. *LO* local oscillator. Channel selection and conversion to a second, lower intermediate frequency (about 120 MHz) take place in the indoor unit *IU*, which is located close to the television set *TV*. Automatic gain control (*AGC*) and limiting are then followed by frequency demodulation (*FDem*). Since most television sets do not have direct video and audio inputs, it is now necessary to amplitude-modulate (*AM*) on the vestigial-sideband system. A phase-locked loop is included to provide frequency tracking of the oscillator used in the frequency demodulation.

The final objective is of course a sufficiently noise-free television picture. A quantity that will indicate whether this has been achieved is the ratio C/N of the carrier power to the noise power for a single television channel (bandwidth 27 MHz). For successful noise-free FM demodulation C/N should be at least 25 (a difference in level of 14 dB).

The quantity C/N depends on two parameters at the receiver end: the gain of the receiver antenna and the noise figure of the receiver. As to the latter parameter — this is very dependent on the semiconductor elements used. The designer is of course dependent on what is available, and the possibilities are therefore limited. The mixer with a Schottky-barrier diode at the input of front-end I is an intrinsically less satisfactory input circuit than the low-noise preamplifier with GaAs field-effect transistors of front-end II; this can be seen from the different noise figures of the two front-ends. In both front-end circuits careful attention is paid to the design of components such as the filters, and an effort is made to reduce signal power losses to a few tenths of a dB.

The other parameter is the gain of the antenna. This is a result of the directional effect. A stronger directional effect, and hence a higher antenna gain, is obtained by using a larger reflector. This approach is very effective: at a diameter of about 1 m a variation of 10 cm gives a difference of 1 dB. The technical means for obtaining the desired signal-to-noise ratio are therefore available; the limitations are essentially of a commercial and practical nature. The manufacturing costs of the antenna dish increase rapidly with increasing diameter; the installation costs also increase, since a larger antenna has to be pointed more accurately at the satellite and the alignment must remain unaltered even in high winds. If the

Table II. Link budget for the transmitter-receiver link, drawn up for the experiments with the Japanese experimental satellite (0 dB corresponds to 1 W). A ratio $C/N \geq 14$ dB is necessary for good noise-free frequency demodulation; both front-ends meet this criterion under favourable conditions, front-end II (with GaAs FETs) even under unfavourable conditions.

	Front-end I	Front-end II
Satellite		
Power of travelling-wave tube	20 dB	20 dB
Gain of transmitting antenna	40	40
Coupling losses	− 1.7	− 1.7
Equivalent isotropically radiated power (EIRP)	58.3	58.3
Space		
Attenuation in space	−205.8	−205.8
Receiver		
Gain of receiving antenna dia. 1 m	39.9	39.9
Carrier power (C) at the input in clear weather	−107.6	−107.6
Equivalent noise power (N)	−124.5	−127.5
C/N in clear weather	16.9	19.9
Variable losses		
Attenuation due to rain	− 4	− 4
Receiver antenna alignment error	− 0.5	− 0.5
Attenuation in troposphere	− 1	− 1
Minimum value of C/N	11.4	14.4

[1] P. Baudet, M. Binet and D. Boccon-Gibod, this issue, p. 269.

diameter is set at 1 m, the value chosen for our designs, the dish can be manufactured in one piece from expanded polyurethane coated with copper foil. The half-power beam width is 1.75°.

The noise performance of the antenna-receiver combination is conventionally expressed by the quantity G/T, where G is the gain of the antenna and T is the 'noise temperature' of the receiver. At WARC '77 it was agreed that for domestic receivers for satellite television G/T should be at least 6 decibels per kelvin. For community antenna systems G/T must be 14 dB/K to compensate for the losses in the distribution system. A larger antenna dish is then necessary.

The 'noise temperature' of the receiver is the temperature that an imaginary resistance at the input of the receiver would have to have to generate the same amount of thermal noise as is actually present in the system. It is assumed in this imaginary situation that the rest of the system is noise-free. The noise power available from the resistance is equal to kTB, where k is Boltzmann's constant, T the absolute temperature of the imaginary resistance and B the bandwidth of the system. The noise power is thus proportional to the absolute temperature. The signal power, on the other hand, is proportional to the antenna gain G, so that the ratio G/T defines a signal-to-noise ratio. If for example T increases by a factor of two, then G must become twice as large to maintain the same signal-to-noise ratio. It is usual — although misleading — to express the ratio G/T in decibels per kelvin; if T is doubled it is *not* the number of decibels that is to be doubled, of course.

Front-end with mixer at the input (front-end I)

Low-noise amplification of signals at the very high frequencies we are concerned with here has only recently become a practical possibility. The conventional solution to the problem (as in radar, for example) has always been to convert at the input to the receiver, by mixing with a locally generated signal, to a lower frequency at which amplification was feasible. The obvious nonlinear element to use for the mixing process is a semiconductor diode with a Schottky barrier. This barrier is located at the transition from a metal to the semiconductor, and unlike the barrier layer in a *P-N* junction, it does not occupy any physical volume of the semiconductor, so that space-charge and other delay effects are small and this kind of diode is particularly suitable for high frequencies.

Even though low-noise preamplification with GaAs FETs is now possible there is a continuing interest in the front-end with a mixer directly at the input; also, with this simpler design it is possible to achieve a value of 6 dB/K for G/T in production with an antenna diameter of 1 m, and this meets the specification put forward at WARC '77.

We shall now look more closely at the technical aspects of front-end I.

Mixer

The mixer is made on a microwave integrated circuit using microstrip lines [2]. It is completed by a two-stage intermediate-frequency amplifier on a single alumina substrate (*fig. 7*).

The signal arriving from the waveguide is fed to the Schottky-barrier diode through a number of filters. These are included

— to exclude image-frequency signals at about 10 GHz;

— to prevent energy from the 11 GHz oscillator from being radiated;

— to reflect mixing products at other frequencies in the correct phase, so that they will contribute to the output power at the intermediate frequency after mul-

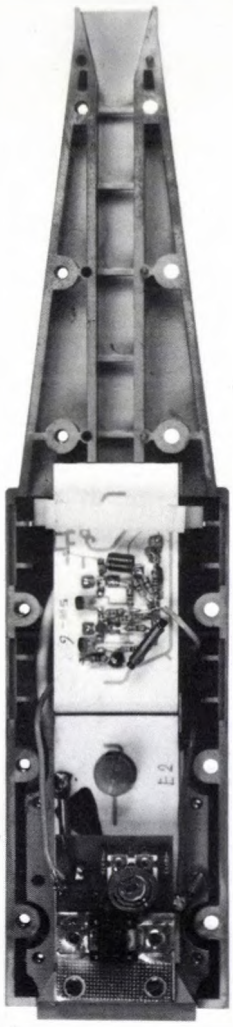


Fig. 7. The 12 GHz converter unit of front-end I, shown opened. The housing is made of metallized plastic. The input horn is shown at the top, with a waveguide below it, containing a 12 GHz bandpass filter formed from two irises. The two white rectangles are microwave ICs (on alumina substrate); the upper IC contains the input mixer and a two-stage intermediate-frequency amplifier, the lower one contains the 11 GHz local Gunn oscillator. The oscillator circuit is partly hidden from view by a unit for d.c. conversion of the supply voltages.

tiple mixing and reflection ('image enhancement').

We should perhaps explain image enhancement a little further. In addition to the fundamental at 11 GHz the local oscillator signal contains a strong second harmonic at 22 GHz (*fig. 8*). When this is mixed with the sum signal at 23 GHz the intermediate frequency at 1 GHz is again produced. If this is in phase with the difference signal of 12 GHz and 11 GHz, there is enhancement. There is also a difference signal at 10 GHz arising from the mixing between 12 GHz and 22 GHz, and this in turn can produce a contribution at 1 GHz after mixing with the signal at 11 GHz.

One of the set of filters is a bandpass filter. This is included in the waveguide as an 'iris filter' (the irises can be seen in fig. 7); the filter limits the received band to 500 MHz around 12 GHz. Other filters are mounted on the substrate; they are partly hidden from view in fig. 7 by a plastic strip.

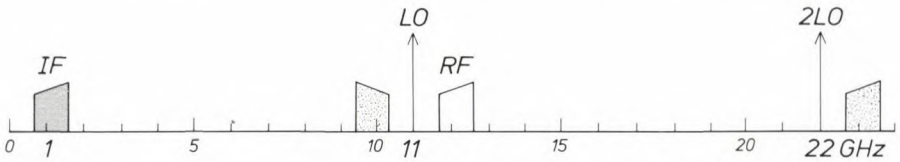
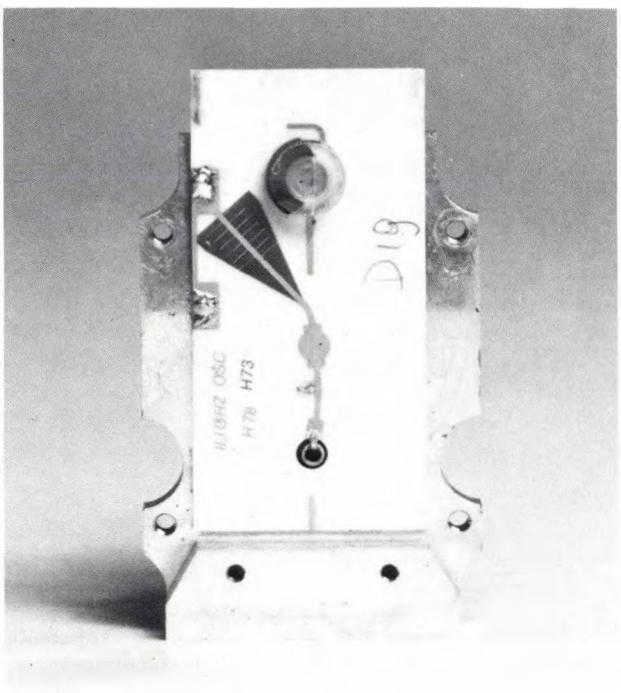
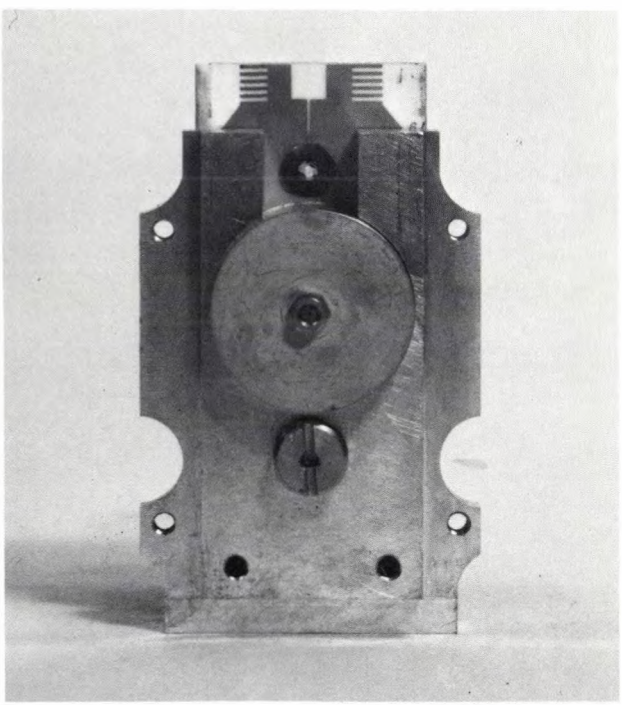


Fig. 8. Frequency spectrum. *RF* received signal band 11.7-12.5 GHz. *LO* local-oscillator frequency. *2LO* second harmonic of the local oscillator. *IF* intermediate-frequency band. The signals produced in the bands around 10 GHz and 23 GHz as a result of the mixing process are reflected back to the mixer by the filters, and after multiple mixing and reflection they make a further energy contribution in the intermediate-frequency band around 1 GHz ('image enhancement').





a *b*

Fig. 9. The 11 GHz Gunn oscillator for front-end I. The Gunn diode is located at the end of a coaxial line. *a*) The other end of the coaxial line can be seen near the bottom of the unit, where it is connected to a microstrip line leading to a 50 Ω matched load (the dark triangle) via an element for coupling to a cavity. The signal coupled out from the cavity is fed to the output through a variable attenuator. *b*) The cavity is soldered directly to the back of the substrate.

Oscillator for 11 GHz

The local oscillator for the first mixer is required to maintain a frequency of 10.8 GHz ± 5 MHz within a temperature range of −25 °C to +55 °C. Its output power should be 10 mW.

[2] J. H. C. van Heuven and A. G. van Nie, Microwave integrated circuits, Philips tech. Rev. **32**, 292-304, 1971.

In front-end I the local oscillator signal is provided by a Gunn diode [3]. To make the frequency of the oscillator thermally stable a cavity is used as the resonant element; variations in temperature do not affect the dimensions of this cavity since it is made of gold-plated Invar. The coupling to the cavity is provided by a microstrip transition designed in such a way that at frequencies other than the desired one the Gunn diode is not coupled to the cavity but to a 50 Ω matched load (the dark triangle in *fig. 9a*; Kurokawa's method). The cavity is soldered directly to the back of the substrate (fig. 9*b*). A variable attenuator connects the cavity to the mixer.

Front-end with 12 GHz preamplifier (front-end II)

The mobility of the charge carriers in GaAs is five times the mobility of the charge carriers in Si; this is one of the reasons why GaAs is a particularly suitable semiconductor for the manufacture of microwave transistors. GaAs MESFETs give the best microwave performance; they are superior to the bipolar transistors above 5 GHz in both noise figure and gain. (MESFET is an acronym for MEtal Semiconductor Field Effect Transistor — a field-effect transistor with a metal gate separated from the bulk of the semiconductor by a Schottky barrier.)

The GaAs MESFETs developed at LEP are used in

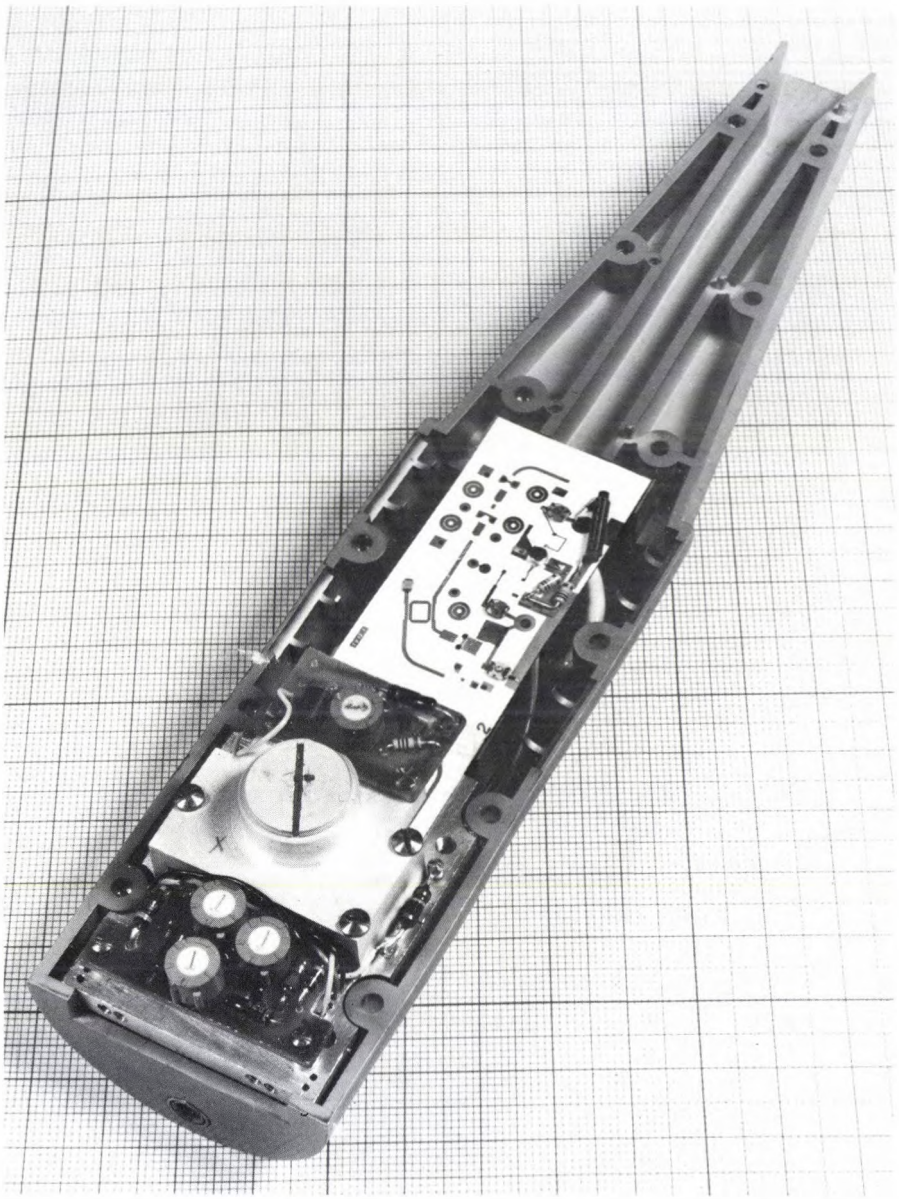


Fig. 10. The 12 GHz converter unit of front-end II, shown opened. The microwave IC that can be seen in the picture contains the two-stage preamplifier with GaAs FETs, the mixer and an intermediate-frequency amplifier. The local oscillator, which also has a GaAs FET, is hidden by a metal cover. A large screw in the cover is used for tuning the dielectric resonator that stabilizes the frequency of the oscillator against temperature changes.

three functions in front-end II [4]. In a two-stage preamplifier advantage has been taken of their low noise figure at 12 GHz (about 2 dB), in the mixer they give a good conversion gain (5.5 dB) and in the oscillator they give a respectable efficiency (18%). Since there is preamplification and conversion gain in the mixer only a single-stage amplifier is necessary at the 1 GHz intermediate frequency. The transistor in this stage is a conventional bipolar high-frequency type, and gives a gain of 13 dB in this circuit.

Preamplifier and mixer

The preamplifier, the mixer and the intermediate-frequency amplifiers are all contained within a single integrated microwave circuit (*fig. 10*). The three GaAs MESFETs (two for the preamplifier and one for the mixer) are mounted directly on the gold metallizing of the alumina substrate. The source of each MESFET is connected to the other side of the substrate, i.e. to earth, by two plated holes. For minimum noise figure a certain optimum impedance must be presented to the gate of the transistor. The impedance-transforming circuit consists of two cascaded lengths of line, a narrow high-impedance line next to the transistor, followed by a quarter-wave line of low impedance (*fig. 11*). A similar cascaded circuit matches the drain of the transistor to 50 Ω, the characteristic impedance of the microstrip line. Circular elements, which take up little space, are used as r.f. decoupling filters in the d.c. supply leads to the transistors [5].

The preamplifier is a broad-band amplifier. In the band from 11 GHz to 12.5 GHz the noise figure is never greater than 3.5 dB and the gain never less than 12.0 dB.

A spur-line filter (fig. 11) serves as a bandstop filter between the preamplifier and the mixer. It prevents signals at the image frequencies around 10 GHz from getting into the mixer (attenuation 40 dB). It also reflects the unwanted mixing products back towards the mixer in the appropriate phase so that the power contained in these mixing products is not lost ('image enhancement').

The 12 GHz signal that passes through the spur-line filter is applied to the gate of the third GaAs MESFET, where the mixing takes place. The 11 GHz signal from the local oscillator is coupled to this line via a directional coupler.

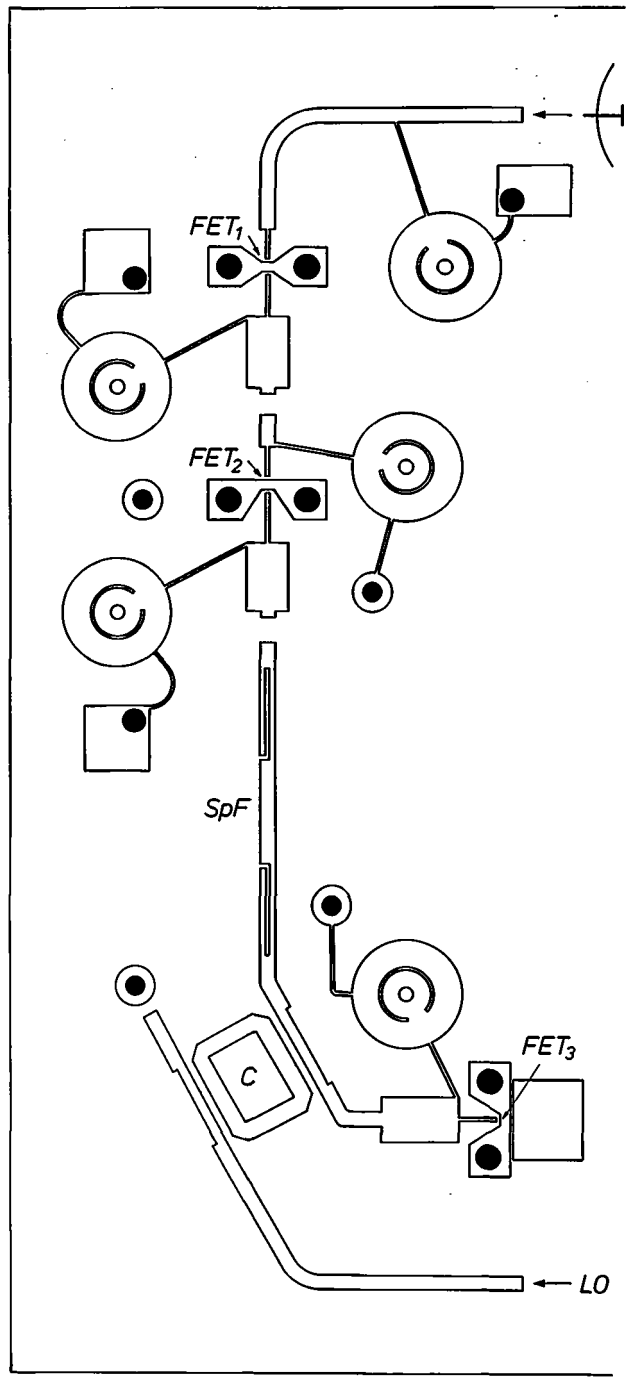


Fig. 11. Part of the pattern of microstrip lines on the first microwave IC of front-end II. The signal from the antenna enters at the top and is applied to the gate of the first GaAs FET (FET_1). The transistor is mounted on a butterfly-shaped pad, with its source directly connected to the pad, which in turn is connected to the back of the IC — i.e. to earth — by two metallized holes. The gate is connected via a bonded gold wire to a narrow high-impedance line to give the correct impedance match. There is a similar arrangement for the drain on the other side. D.C. voltages are applied through decoupling filters consisting of lengths of high-impedance line in series with circular structures that behave like short-circuits at the microwave frequencies. After a second FET amplifier stage (FET_2) there is a spur-line filter *SpF*, which acts as an image-frequency stop filter. The 11 GHz signal from the local oscillator *LO* is coupled in by the directional coupler *C* and both signals are fed to the gate of the third GaAs FET (FET_3), where the mixing takes place.

[3] G. A. Acket, R. Tijburg and P. J. de Waard, The Gunn diode, Philips tech. Rev. 32, 370-379, 1971.

[4] R. Dessert, P. Harrop, B. Kramer and T. Vlek, All FET front-end for 12 GHz satellite broadcasting reception, Proc. 8th Eur. Microwave Conf., Paris 1978, pp. 638-644.

[5] These circular elements have been designed by Ir F. C. de Ronde, now with LEP.

Local oscillator for 11 GHz

A GaAs MESFET oscillator [6] has the advantage of a higher efficiency compared with a Gunn diode (18% instead of 2 to 3%). Another advantage is that the power from the oscillator can easily be varied by varying the drain voltage.

The required temperature stability for the frequency has been obtained in front-end II by tuning the local oscillator with a dielectric resonator — a cylinder of dielectric with a high dielectric constant provides the coupling between two microstrip lines (*fig. 12*). The dielectric material used for this cylinder is barium titanate, which has a relative permittivity of 40. Its temperature coefficient is small and positive and is compensated by a small negative temperature coefficient in other components, giving an excellent temperature stability: ± 1 MHz from −25 °C to +55 °C. The local oscillator can be mechanically tuned through a band of 50 MHz by the screw above the resonator.

Comparison of front-ends I and II

Table III gives a comparison of the front-ends I and II. The values quoted are taken from the Japanese experiments. The carrier-to-noise ratio C/N was also measured in these experiments for both front-ends, and *fig. 13* shows the percentage distribution over the time for the various measured C/N ratios for the month of June 1979. The low C/N levels measured during a small percentage of the time are associated with unfavourable atmospheric conditions.

The front-end with GaAs MESFETs gives the better noise performance. This is mainly of interest for receiving foreign programmes near the edges of the 'footprints'. At other locations it would be possible to take advantage of the better noise performance by making the antenna smaller and so compensating for the higher costs of this receiver with a cheaper antenna that is more easily aligned.

A combination of elements from the two designs is also a likely possibility. It would for example be possible to combine a mixer with Schottky diodes with an oscillator using a GaAs MESFET, with the object of producing a cheaper version. Another possible way of reducing the cost could be to combine the various GaAs elements in a single monolithic integrated circuit. It will not be long before the configuration is decided for the receivers that will pick up the first experimental transmissions from the French and German television satellite in 1983.

[6] P. Lesartre, R. Dessert and J. Magarshack, Stable FET local oscillator at 11 GHz with electronic amplitude control, Proc. 8th Eur. Microwave Conf., Paris 1978, pp. 264-268.

Table III. Comparison of the front-ends I and II in the Japanese experiment.

	Front-end I	Front-end II
Received frequency band	11.9-12.3 GHz	11.9-12.3 GHz
Intermediate-frequency band	900-1300 MHz	900-1300 MHz
Noise figure	6.3 dB	4.0 dB
Gain of frequency converter	30 dB	42 dB
Frequency stability −25 to +55 °C	± 5 MHz	± 1 MHz
Attenuation at image frequencies	60 dB	60 dB
Power supply required	−15 V, 350 mA	−15 V, 50 mA

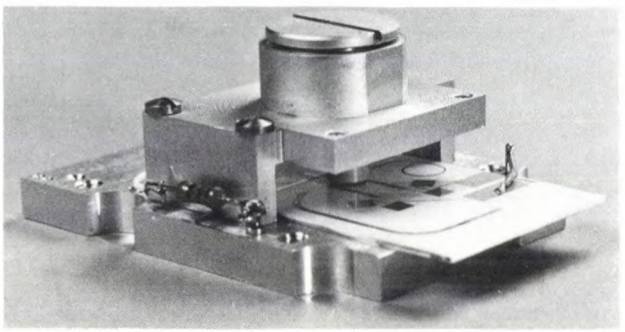


Fig. 12. The 11 GHz local oscillator of front-end II. Underneath the adjusting screw there is a cylinder of barium titanate that provides the coupling between two microstrip lines on the microwave IC. The cylinder is a dielectric resonator and is virtually unaffected by changes in temperature, so that the oscillator frequency does not vary by more than 1 MHz between −25 °C and +55 °C. The tuning screw gives a range of 50 MHz.

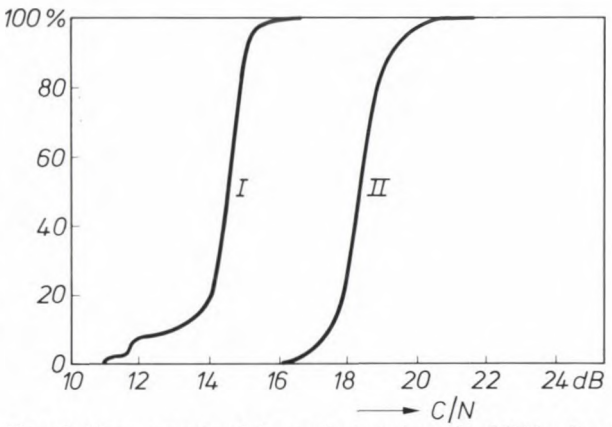


Fig. 13. Measurements of the carrier-to-noise ratio C/N for front-ends I and II during the month of June 1979 in Japan. The curves show the percentage of the time for which a particular level is not exceeded. For a small percentage of the time the C/N levels are low because of effects such as heavy rainfall. Carrier and noise levels were measured after the second frequency conversion. The antenna diameter was 1.0 m.

Summary. Both the French and the West Germans plan to launch a geostationary satellite for television broadcasting in 1983. For the reception of the 12 GHz signal two receiver front-ends have been developed that can be combined with a conventional television set. Both front-ends consist of a unit located at the focus of a 1 m diameter parabolic reflector and producing a frequency conversion to 1 GHz, and a second frequency converter plus FM/AM converter, located indoors. At the input of one of the front-ends there is a mixer with a Schottky-barrier diode and a local oscillator with a Gunn diode; the noise figure in the intermediate-frequency band is 6.3 dB. The other front-end has a two-stage preamplifier, a mixer and a local oscillator, with each stage based on a low-noise GaAs MESFET developed at LEP; the overall noise figure is 4 dB.

Low-noise microwave GaAs field-effect transistor

P. Baudet, M. Binet and D. Boccon-Gibod

Elsewhere in this issue two alternative designs are described for a 12 GHz receiver front-end intended for the reception of television satellite broadcasts. One of the designs uses GaAs field-effect transistors throughout — in the preamplifier and in the mixer stage, and as a local oscillator. It is based on the low-noise GaAs FETs developed by the present authors.

Introduction

For many years gallium arsenide (GaAs) has been preferred to silicon (Si) as a semiconductor material for microwave devices, because of its considerably higher electron mobility, and GaAs field-effect transistors have been made for this application. In field-effect transistors the crucial parameter is the gate length; the gate should be as short as possible, for one reason to ensure that the transit time for the carriers in the channel is short. We have succeeded in reducing the gate length to submicron size (see *fig. 1*) by means of a self-alignment technique. This short gate length gives an excellent high-frequency performance: at 12 GHz the minimum noise figure is 1.9 dB with an associated gain of 7.8 dB.

Our GaAs field-effect transistors are MESFETs (MEtal-Semiconductor Field-Effect Transistors): the gate metal is directly applied to the semiconductor surface to form a Schottky-barrier diode, which has very good high-frequency characteristics.

GaAs MESFETs

History

The idea of controlling the current in a semiconductor by the electric field from a nearby electrode had already been patented by J. E. Lilienfeld in the thirties. But it was only in the fifties, while modern semiconductor technology and photolithographic techniques were developing, that the field effect became more than a laboratory curiosity. In 1952 W. Shockley developed the theory of what he called the 'field-effect transistor', in which the early metallic electrode was replaced by a *P-N* junction made directly at the surface of the semiconductor [1]. In his published article Shockley shows that field-effect transistors can have a better high-frequency response than bipolar transistors of comparable dimensions. Later, in 1966, C. A. Mead [2] of the California Institute of Technology proposed that the *P-N* junction should be replaced by a metal-semiconductor junction. He made an active device with GaAs that met a particular need at the time because of the difficulty of forming good *P-N* junctions in a high-band-gap material like GaAs. Transistors of this type are now called MESFETs.

Since then significant progress has been achieved and GaAs MESFETs have given excellent low-noise and high-gain performance at microwave frequencies. Today the GaAs MESFETs surpass the best bipolar transistors above 5 GHz and have lower noise figure, higher power gain, higher output power and a better power efficiency. The GaAs MESFETs we have produced in our laboratories can be used at frequencies up to 12 GHz.

Principle of operation

A cross-section of a GaAs MESFET structure is shown in *fig. 2*. Current conduction takes place in a thin GaAs *N*-type layer appropriately doped and grown on a high-resistivity GaAs substrate. Two ohmic contacts, the source and the drain, are provided on each side of a metallic gate electrode applied to the semiconductor surface and forming a Schottky-

P. Baudet, doct. 3me cycle, M. Binet, doct. 3me cycle, and D. Boccon-Gibod, doct. 3me cycle, are with Laboratoires d'Electronique et de Physique Appliquée (LEP), Limeil-Brévannes (Val-de-Marne), France.

[1] W. Shockley, A unipolar "field-effect" transistor, Proc. I.R.E. 40, 1365-1376, 1952.
[2] C. A. Mead, Schottky barrier gate field effect transistor, Proc. IEEE 54, 307-308, 1966.

barrier diode with it. The conductance of the channel between source and drain is modulated by applying a transverse electric field obtained by biasing the gate with respect to the substrate (which is connected to the source). As a result, there is a depletion layer beneath the gate, and this layer widens when the applied voltage between the gate and the channel increases. The conducting channel becomes narrower by the same amount and the current that can flow in the channel is therefore decreased. The operation is different from that of an MOS field-effect transistor [3], in which the biased gate attracts minority carriers to form an inverted layer directly beneath the gate; it is this inverted layer that conducts the current.

In the GaAs field-effect transistor current conduction can take place under two distinct regimes. The

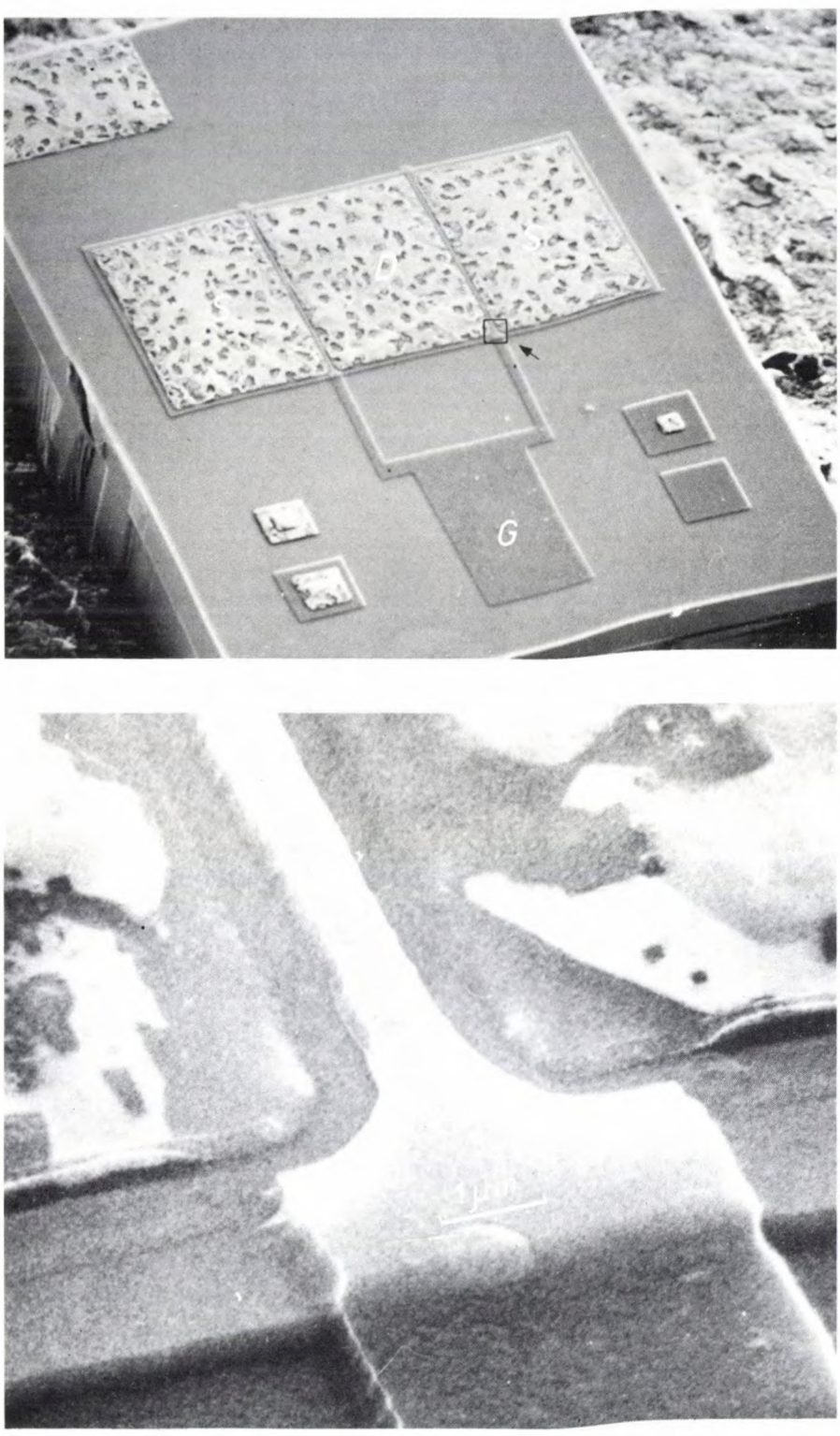


Fig. 1. Microwave GaAs field-effect transistor. *a*) The complete transistor. *G* gate pad, connected to the two 'fingers' of the gate. *S* source. *D* drain. *b*) Enlarged view of one of the 'fingers' of the gate where it passes over the edge of the epitaxial active *N*-layer.

case of low drain and gate bias is illustrated in *fig. 3a*. If the drain-to-source voltage V_{DS} only differs slightly from zero, the voltage at each point x of the channel increases uniformly from 0 V at the source to V_{DS} at the drain; there is a progressive extension of the depleted region. Since the current I_{DS} is the same everywhere in the channel, this effect is compensated by a progressive increase in the velocity $v(x)$ of the carriers, i.e. the electrons, from source to drain. This requires a corresponding progressive increase in the electric field-strength $E(x)$, since the mobility μ_0 of the electrons in the relation

$$v(x) = \mu_0 E(x) \qquad (1)$$

is assumed to be constant.

When the drain bias increases beyond $V_{DS\,SAT}$ the current conduction enters into the second regime (fig. 3*b*). The electric field reaches a critical value E_0; when it exceeds E_0 the drift velocity of the electrons, and hence the drain current, no longer increase. There are therefore two distinct zones beneath the gate: a zone on the source side where the electron mobility can be considered as a constant, and a zone on the drain side where the electron velocity has reached a saturation level.

In reality the analysis is somewhat more complicated than this, because the drift velocity in GaAs goes through a peak as the field-strength increases and then falls to a saturation value. This does not affect the general principles illustrated in fig. 3, however.

Advantages of GaAs over Si

The amplification to be obtained from a field-effect transistor is proportional to the change δI_{DS} in the drain current due to a given change δV_{GS} in the gate voltage. The ratio is called the transconductance:

$$g_m = \frac{\partial I_{DS}}{\partial V_{GS}}.$$

To a first approximation, the transconductance for the operation of the transistor (i.e. neglecting the parasitic elements) when the current just starts to saturate, is given by

$$g_m = k_1 \frac{\mu_0 N_D}{l}, \qquad (2)$$

where N_D denotes the doping level of the channel and l the gate length, and k_1 is a constant depending on the geometry and the bias conditions. When the current saturates, the transconductance becomes proportional to the limiting velocity v_s alone.

The electron mobility μ_0, the doping level N_D and the limiting velocity v_s are properties of the material; the values for GaAs are clearly superior to those for

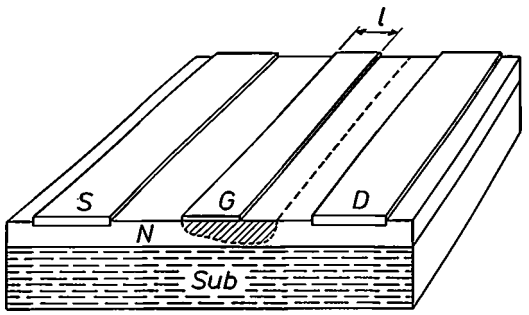
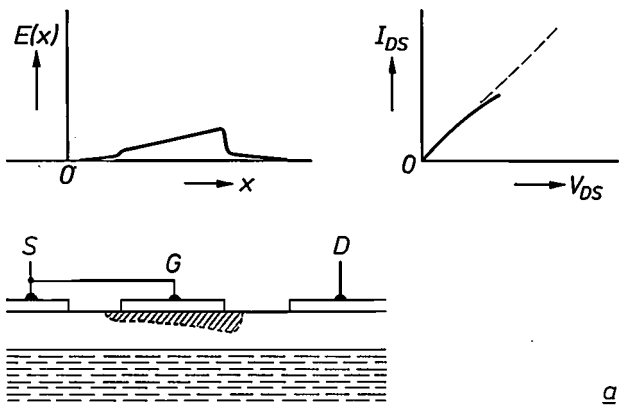


Fig. 2. Schematic cross-section of GaAs MESFET. Source (*S*) and drain (*D*) form ohmic contacts with the active *N*-type GaAs layer (*N*); the gate (*G*) is insulated from it by a reverse-biased Schottky-barrier diode. The depletion layer beneath the gate restricts the electric current path to a narrow channel above the high-resistivity GaAs substrate *Sub*.



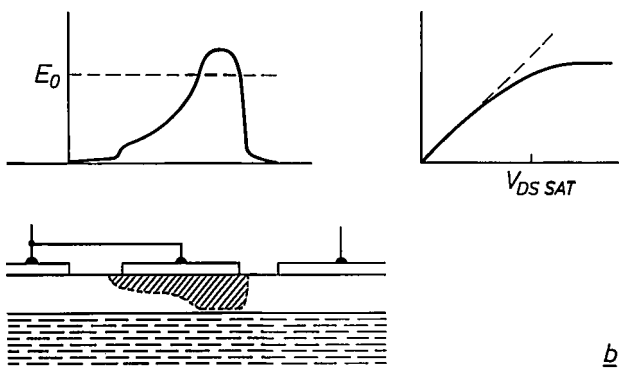


Fig. 3. *a*) Non-saturated regime. The gradual narrowing of the cross-section of the conducting channel from source to drain is made up for by an increasing carrier velocity due to a gradually increasing electric field $E(x)$; x distance from source. The drain current I_{DS} increases linearly with the drain voltage V_{DS}. *b*) Saturated regime. Above a certain value E_0 an increase of the electric field-strength no longer results in a higher carrier velocity. Above $V_{DS\,SAT}$, a higher drain voltage does not therefore give a greater drain current. The gate depletion region leaves a channel for the charge carriers that corresponds exactly to the current and the saturated carrier velocity.

[3] J. A. van Nielen, Operation and d.c. behaviour of MOS transistors, Philips tech. Rev. 31, 209-215, 1970.

Si: the electrons have a mobility six times as high and the maximum drift velocity is twice as high as in Si (*Table I*). This contributes greatly to a higher transconductance. The energy gap is also higher, permitting a higher doping level (up to $2.5 \times 10^{17}/cm^3$) for the same breakdown voltage, which helps to keep the parasitic resistances low.

Equation (2) shows that for the non-saturated transistor the transconductance is inversely proportional to the gate length l. This is illustrated by *fig. 4*. A short gate length is therefore desirable; in addition, this shortens the electron transit time beneath the gate, thus improving the high-frequency performance. In developing our GaAs MESFETs, making a short gate was one of the main items we concentrated our effort on. The 0.5 μm gate that resulted is close to the present limit of the technology.

Table I. Comparison of physical properties of Si and GaAs, both with N-type doping at 10^{17} atoms per cm^3.

	Si	GaAs
Electron mobility	800	4700 cm^2/Vs
Saturated electron velocity	0.9×10^5	1.8×10^5 m/s
Energy band gap	1.1	1.4 eV

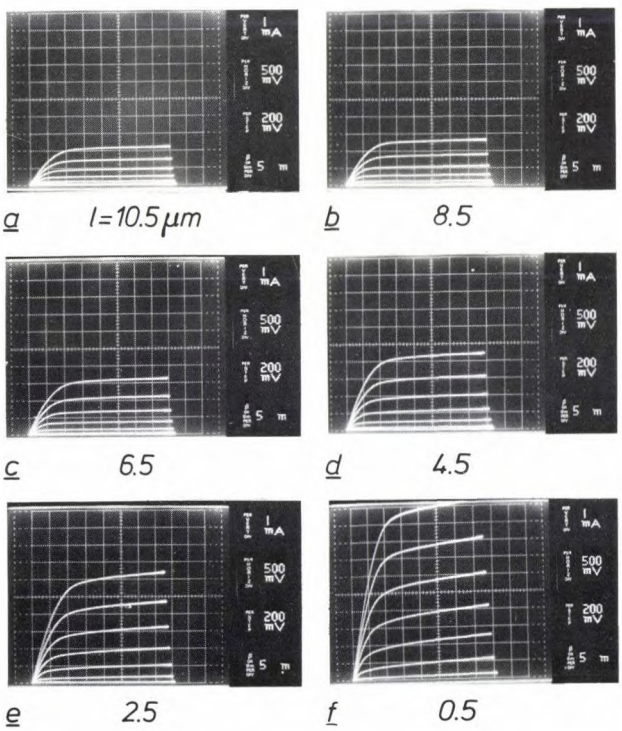


Fig. 4. Sets of characteristic I_{DS}-V_{DS} curves for GaAs MESFETs of different gate lengths l from 10.5 μm (*a*) to 0.5 μm (*f*). The transconductance, which is indicated by the spacing of the curves in each set, increases as the gate length decreases. Vertical scale: 1 mA per square; horizontal scale: 500 mV per square; the gate voltage varies in steps of 200 mV between curves.

High-frequency limitations

There are several high-frequency limitations which, for a given material, are all dependent on device geometry. In particular, high-frequency limits are imposed by the transit time of the electrons beneath the gate and by the time required to charge and discharge the gate capacitance C_{gs}. This transit time is represented by the time constant of the input network of the intrinsic transistor, which can be assumed to consist of C_{gs} in series with an equivalent channel resistance R_{gs}. In practice, the most severe frequency limitation is described by the expressions for the current gain $G_c = \partial I_{DS}/\partial I_{GS}$, which clearly shows the dependence on geometrical parameters. As a first approximation

$$G_c = \frac{g_m}{\omega C_{gs}}, \qquad (3)$$

where $\omega = 2\pi \times$ frequency. The current gain decreases with increasing frequency and reaches the value 1 at the transition frequency

$$f_T = \frac{g_m}{2\pi C_{GS}}. \qquad (4)$$

To make f_T as high as possible g_m is made high and C_{gs} small. Both objectives are reached by making the channel as short as possible. This underlines what has been said above about the importance of a short channel. For our GaAs MESFET the transition frequency is 12.9 GHz.

Obviously, the other dimensions have to be reduced as well to minimize the parasitic elements such as the equivalent channel resistance, which determines the input time constant.

Noise in GaAs MESFETs

In the intrinsic field-effect transistor the thermal agitation of the electrons in the channel is the main noise source [4]. Generation-recombination noise has a low-frequency character and can be neglected at microwave frequencies. Additional noise generation due to electron interactions with lattice phonons, scattering of the electrons between two potential valleys in the conduction band, and high-field carrier diffusion should, however, be included. The thermal agitation of the electrons in the channel causes random fluctuations of the drift velocity and hence of the drain current. As the channel region is capacitively coupled to the gate, the thermal agitation of the electrons in the channel will also induce random gate currents, which are a second source of noise. This 'gate noise', however, is partly correlated with the channel noise and has a tendency to reduce it, since it has the opposite phase. The absolute value of the correlation coeffi-

cient c is about 0.3 or 0.4; the effect of the correlation is not very great, as can be seen from an approximate expression for the minimum noise figure of the intrinsic transistor:

$$F_{min} \approx 1 + 2\sqrt{1-c^2}\,\frac{f}{f_T} + \ldots . \qquad (5)$$

Higher-order terms in f/f_T are omitted from the expression as they are of minor importance for operation below f_T.

The intrinsic transistor is connected to the outside world through parasitic resistances. In our transistor the resistances in series with the source and drain both have a value of 0.5 ohm, and the gate resistance is 2 ohms; the higher value is related to the narrow gate metallization. The parasitic resistances contribute considerably to the degradation of the noise figure.

Device technology

Preparation of the active layer

The N-type active layer is grown on a GaAs substrate made semi-insulating (10^8 ohm cm) by a chromium dopant. The epitaxial growth of the active layer takes place from the gaseous phase ($AsCl_3/Ga/H_2$ system). The thickness of the electrically active layer is 0.2 μm.

When layers are fabricated in this way, the interface with the substrate tends to have poor electron mobility (*fig. 5a*). This effect is related to the quality of the substrate, and even more closely to the starting-growth conditions of the active layer. In operation, the carriers beneath the gate are constrained to this interface region and consequently the gain and noise performance are poor. The situation is corrected by keeping the carriers away from the interface by means of a high-resistivity, weakly doped buffer layer (*BL* in fig. 5*b*). As a result, the electron mobility increases from the active layer towards the buffer layer while the doping level decreases. The buffer layer and the active layer are grown in the same process.

Gate definition by self-alignment

The submicron gate length required in low-noise GaAs MESFETs is not easily obtained by conventional methods such as the standard photolithographic techniques based on the alignment of a sequence of optically generated photomasks. Some sort of self-alignment is necessary. Whereas other methods attain self-alignment of the gate with the source and drain contacts by using the undercutting or 'overflowing' of ohmic contacts, we have developed a method in which the gate pattern is used to align the source and drain contacts [5].

The process consists of the following stages.

1. The 0.2 μm thick active N-type layer is divided into active zones for the individual transistors by chemical etching to a depth of 0.4 μm. During this etching the wafer is oriented in such a way that sloping 'mesa' edges are obtained because of the anisotropic behaviour of GaAs; the metallization can easily pass over these edges continuously.

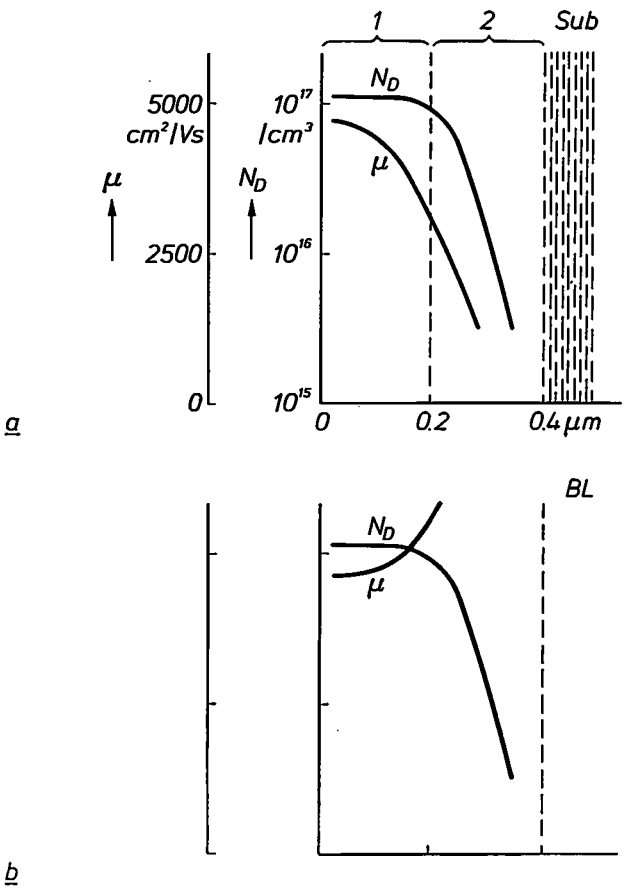


Fig. 5. *a*) The electron-mobility μ in the N-layer tends to decrease towards the interface with the semi-insulating substrate *Sub*, which lies at a depth of 0.4 μm from the surface. There is a transition layer *2* between the active layer *1* and the substrate. *b*) To correct this, a weakly doped high-resistivity buffer layer *BL* is inserted between the N-layer and the substrate. The mobility increases towards the buffer layer.

2. Aluminium is deposited on the freshly deoxidized surface under ultra-high vacuum, thus producing a high-quality Schottky diode (*fig. 6a*). The wafer is next coated with photoresist, which is then removed from the source and drain areas. Aluminium is etched away through the windows in the photoresist until the

[4] P. A. H. Hart and F. M. Klaassen, The MOS transistor as a small-signal amplifier, Philips tech. Rev. **31**, 216-224, 1970.
[5] P. Baudet, M. Binet and D. Boccon-Gibod, Submicrometer self-aligned GaAs MESFET, IEEE Trans. **MTT-24**, 372-376, 1976.

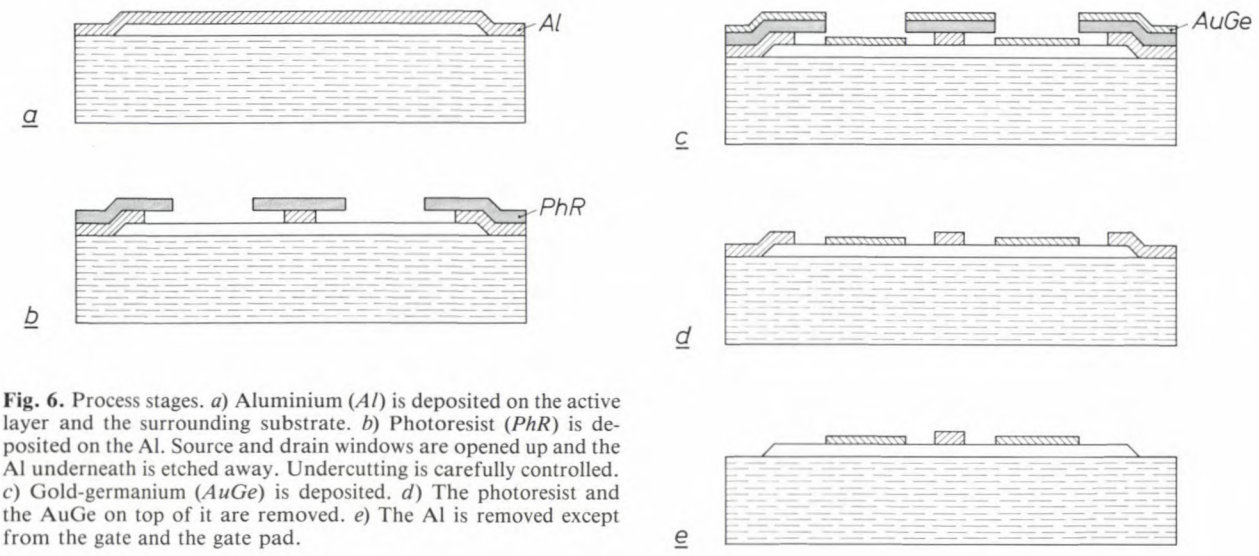


Fig. 6. Process stages. *a*) Aluminium (*Al*) is deposited on the active layer and the surrounding substrate. *b*) Photoresist (*PhR*) is deposited on the Al. Source and drain windows are opened up and the Al underneath is etched away. Undercutting is carefully controlled. *c*) Gold-germanium (*AuGe*) is deposited. *d*) The photoresist and the AuGe on top of it are removed. *e*) The Al is removed except from the gate and the gate pad.

undercutting of the aluminium on either side of the gate contact is a third of the channel length (fig. 6*b*).
3. After the source and drain GaAs areas have been cleaned an AuGe layer is evaporated on to the wafer. The overhang of the photoresist ensures that the contacts are automatically aligned correctly (fig. 6*c*). At other places the AuGe is lifted off after the photoresist has been dissolved (fig. 6*d*).
4. Etching the aluminium again with a final mask defines a gate pad at one end of the channel and the remaining Al is removed (fig. 6*e*). Annealing at 450 °C in a stream of hydrogen makes the drain and source contacts ohmic.

Performance

Gain and noise

Fig. 7 shows the gain and noise performance of a GaAs MESFET with a 0.5×200 μm gate (two 'fingers' of 0.5×100 μm). The gain measurements were made with the source impedance at the input of the transistor adjusted to the value for the minimum noise figure. This impedance value is slightly different from the 'matched' value giving maximum available gain. A low drain-to-source current I_{DS} was used, to give the best compromise between C_{gs} and g_m, so that g_m/C_{gs} was at a maximum and the noise at a minimum.

Equivalent circuit

An equivalent lumped-element circuit of the GaAs MESFET including the parasitics of the mounted transistor is shown in *fig. 8*. Each circuit element corresponds to a physical element. Numerical values are given.

Applications

The gain and noise performances mentioned above are very promising and MESFETs often provide the best method of obtaining low-noise small-signal amplification, e.g. in microwave receivers for radar and telecommunication. Our GaAs MESFETs are now being considered for application in TV tuners.

For amplifiers of small bandwidth, say less than 10%, very simple circuits can be used to match the transistor to the characteristic impedance of the

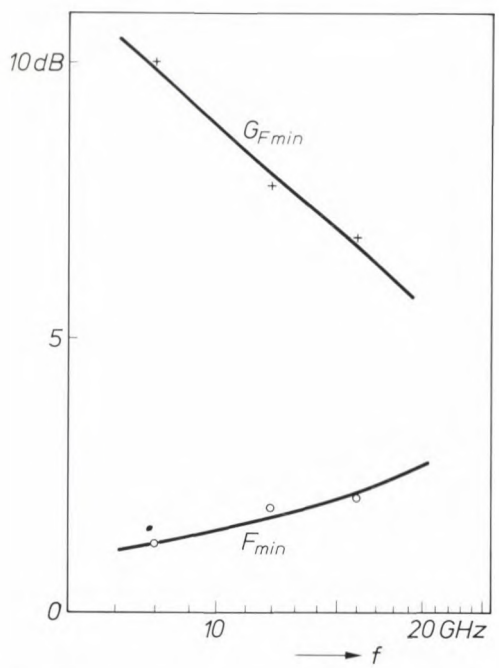


Fig. 7. Minimum-noise gain G_{Fmin} and minimum noise figure F_{min} of one of our 0.5×200 μm gate GaAs MESFETs at three microwave frequencies f. Input impedance and drain current have been selected for minimum noise figure.

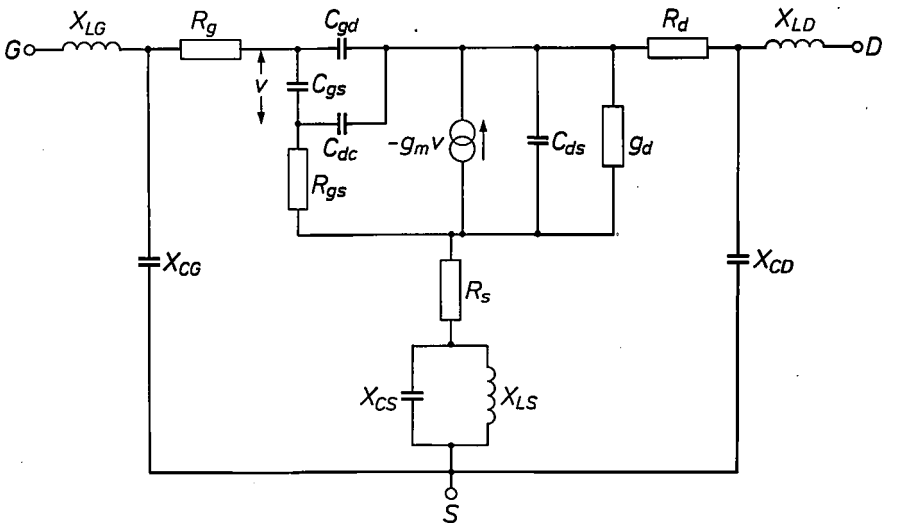


Fig. 8. Equivalent lumped-element circuit of GaAs MESFET. *G*, *S*, *D* gate, source, drain. The source is assumed to be connected to the ground plane of the alumina substrate.

C_{gs}	gate-to-channel capacitance	0.28 pF
C_{gd}	drain-to-gate feedback capacitance	0.007 pF
C_{dc}	drain-to-channel capacitance	0.006 pF
C_{ds}	drain-to-source capacitance	0.09 pF
R_{gs}	channel resistance	8.7 Ω
R_g	gate resistance	2 Ω
R_s	source-to-channel resistance including contact resistance	0.5 Ω
R_d	drain-to-channel resistance including contact resistance	0.5 Ω
g_m	transconductance	22.7 mA/V
g_d	output conductance	0.17 mA/V
τ	delay corresponding to electron transit time in saturated region	4.3 ps
X_{LG}	parasitic inductance of gate bonding wire	0.3 nH
X_{LS}	parasitic inductance of source bonding wire	0.01 nH
X_{LD}	parasitic inductance of drain bonding wire	0.25 nH
X_{CG}	parasitic capacitance between gate pad and earth	0.005 pF
X_{CS}	parasitic capacitance between source pad and earth	0.03 pF
X_{CD}	parasitic capacitance between drain pad and earth	0.02 pF

microstrip line. If a single microstrip line does not give an adequate match, two cascaded impedance-transforming line sections are necessary. The first is generally a narrow high-impedance line (100 ohms), which resonates with the impedance of the transistor, thus cancelling the reactive component of the impedance. The second section is a quarter wavelength of wide low-impedance line which matches the resistive component of the impedance to the 50 ohm line. *Fig. 9* shows the layout of a microstrip circuit for matching the gate and drain of the FET to the 50 ohm lines. The bias voltages are applied to the electrodes through quarter-wave high-impedance transformers designed to keep the high-frequency signals out of the d.c. supply.

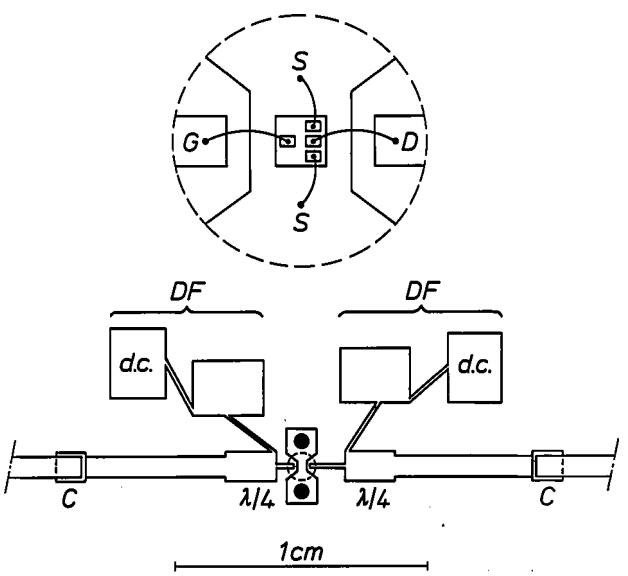


Fig. 9. Layout of microstrip circuit designed to match GaAs MESFET to the line impedance. The method of connecting the MESFET by means of bonding wires is shown. Narrow (high-impedance) line sections of suitable length and wide (low-impedance) line sections of about a quarter-wavelength ($\lambda/4$) are combined to form impedance transformers. *G*, *S*, *D* gate, source, drain. *C* isolating capacitor. *DF* decoupling filters for high-frequency decoupling of the bias-supply bonding pads *d.c.*

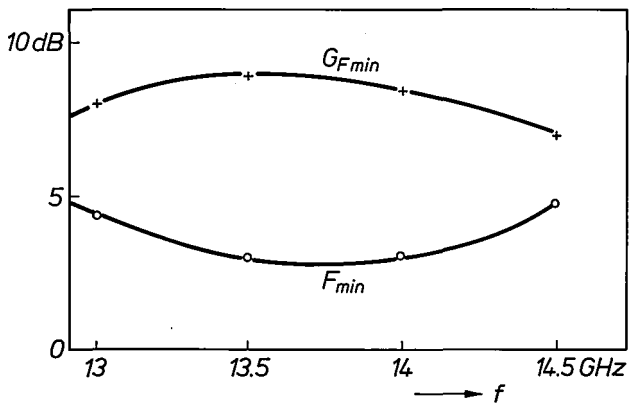


Fig. 10. Minimum-noise gain G_{Fmin} and minimum noise figure F_{min} of microstrip amplifier circuit containing GaAs MESFET, as a function of frequency f.

Fig. 10 shows the performance of an amplifier using this circuit: at 13.75 GHz the noise figure is 2.7 dB and the associated gain 8.6 dB; this gain is maintained to within 1 dB in a band 1.5 GHz wide.

The simple circuit described here is sufficient for most narrow-band specifications. For wide-band applications, more complicated circuits yield satisfactory results after a suitable choice of topology and computer optimization of the parameters.

Summary. LEP have developed a GaAs MESFET with a gate length of 0.5 μm. The correct geometry has been obtained by self-alignment of the source and drain contacts. The microwave performance is excellent: a minimum noise figure of 1.9 dB at 12 GHz and an associated gain of 7.8 dB.

An equipment for measuring the magnetic fields of television deflection coils

G. A. J. van Engelen, J. L. M. Hagen and W. A. L. Heijnemans

In an article on the 30AX colour picture system [1] which appeared earlier in this journal, the design of the deflection coils was described as 'a difficult but fascinating game of chess'. In the article below the authors describe the instrumentation they have devised for the designer — an aid that will keep him informed of the results of every move.

Introduction

One of the most important aspects in the design of a deflection unit for a television receiver is to prevent the occurrence of deflection errors that show up as visible defects in the picture on the screen. Such defects may appear as poor picture focus and as distortion. In a colour picture the convergence may be defective: the red, the green and the blue parts of the image are no longer exactly in register.

To achieve colour convergence over the entire screen it used to be necessary in earlier deflection systems to adjust the convergence at a number of points for every set. In the 30AX system [1], the latest Philips development, introduced early in 1978, the coil design has been refined to such an extent and the accuracy of production so greatly improved that no adjustment is required at all.

In the design of such a deflection system the 'adjustable' coil parameters have to be chosen in such a way for the given tube geometry that a picture containing a minimum of deflection errors appears on the screen. The first question to be answered here is what the spatial distributions of the horizontal and vertical deflection fields 'should look like'. The next question relates to the nature and dimensioning of the coil parameters that are necessary to achieve these optimum deflection fields.

To answer the first question the coil designer must know the characteristics of the picture tube for which the deflection unit is intended, such as the size of the screen, the maximum deflection angle and the structure of the three electron beams. The optimum deflection fields can be determined theoretically and with the aid of numerical analyses, including computer calculation of the paths of the deflected electrons.

Answering the second question is more difficult. Numerical calculations have their limitations here. Because of the complicated geometry of a deflection coil it is impossible to give any detailed description in a manageable model.

This problem in particular made it desirable to be able to measure the deflection fields. At the same time it would be possible to investigate the tolerances and find out which field components are responsible for certain deflection errors.

Principle of the measurement

To understand the behaviour of the magnetic deflection field in the entire deflection region, it is not necessary to fill this region with data points. If it is assumed that the deflection region is free of magnetic sources (i.e. that the region contains no electric currents and no magnetized material), then the deflection field can be derived from a scalar magnetic potential Φ. To determine this it is sufficient to measure the field-strength at suitable points on the boundary of the deflection region, which coincide approximately with the glass wall of the picture tube. From the results of this measurement the potential Φ at these

G. A. J. van Engelen, J. L. M. Hagen and Ir W. A. L. Heijnemans are with Philips Research Laboratories, Eindhoven.

[1] W. A. L. Heijnemans, J. A. M. Nieuwendijk and N. G. Vink, The deflection coils of the 30AX colour picture system, Philips tech. Rev. 39, 154-171, 1980 (No. 6/7).

points can be determined, and from these values it is possible to calculate the potential and the field at all points *in* the deflection region. We use cylindrical coordinates (z,r,ψ), where r is the radius, ψ the angle and z the coordinate along the common axis of the picture tube and the deflection coils. To determine the magnetic potential Φ everywhere along the edge of the deflection region, we always have to measure the field-strength component H_ψ at fixed r and z. This is done by setting out a number of circles along the boundary surface; the plane of each circle is perpendicular to the axis of the coil and the centres all lie on the axis; see *fig. 1a*. The field measurements are now performed for each circle at a number of equidistant points (fig. 1*b*). The measured data is processed by making a Fourier analysis of H_ψ for each circle. Each term of the Fourier series of H_ψ yields, after some calculation, the amplitude of the corresponding multipole component of the deflection field [1].

In the field measurement the absolute field values are not as important as the spatial distribution of the deflection fields. Measurement of the relative field values is therefore sufficient, since we assume — and this is confirmed in practice — that the magnetic field-strength at any given point in the deflection region is linearly related to the current through the deflection coils. We do not therefore need to take any account of saturation or eddy-current effects. In view of this linearity, it is sufficient to perform a complete series of measurements for a single value of the deflection current. When an alternating current is used in the deflection coil and a small coil is used as a field probe, a linear signal is obtained that is insensitive to other fields, e.g. the Earth's magnetic field.

The measuring equipment

The electrical circuits

Fig. 2 shows the electrical block diagram of the measuring equipment, and also indicates the directions in which the probe and the deflection coil move. The horizontal deflection coil is supplied by an alternating current at 5.0 kHz, the vertical coil by an alternating current at 4.1 kHz. The probe C is formed by a small coil (see fig. 3, inset) whose length and diameter (1.5 mm and 2 mm respectively) are such that the induced alternating voltage is determined as far as possible by the magnetic field-strength at the centre.

Depending on the position of the probe, the signal level at the input of the preamplifier A varies from a few μV to 10 mV. The components at 5.0 kHz and 4.1 kHz that constitute this signal are separated from each other by two synchronous detectors D_H and D_V. The measured signals are supplied to a P855 minicomputer via the analog/digital converters A/D.

As indicated in fig. 2, the probe C only moves in the r,z-plane. The deflection unit is mounted vertically on a turntable, which can be displaced in steps of 2.5° in relation to the probe (the ψ movement).

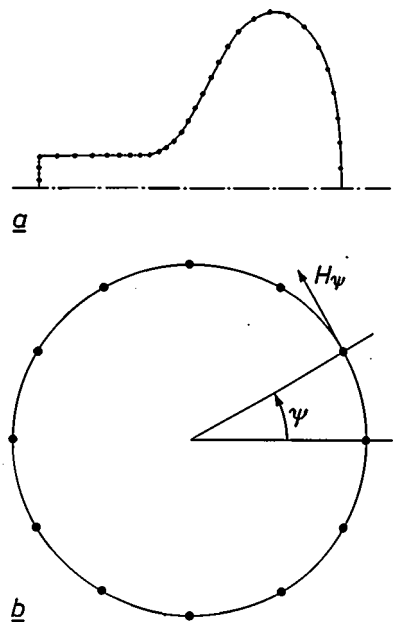


Fig. 1. *a*) Schematic representation of the points of intersection of the measurement circles with a plane through the z-axis. The circles are perpendicular to the z-axis and coincide approximately with the wall of the picture tube. The distribution is most dense where the fastest spatial changes of the deflection fields are to be expected, i.e. at the transition between neck and cone. *b*) For each circle the field component H_ψ, i.e. the component tangential to the circle, is determined at a number of equidistant points (from 8 to 72).

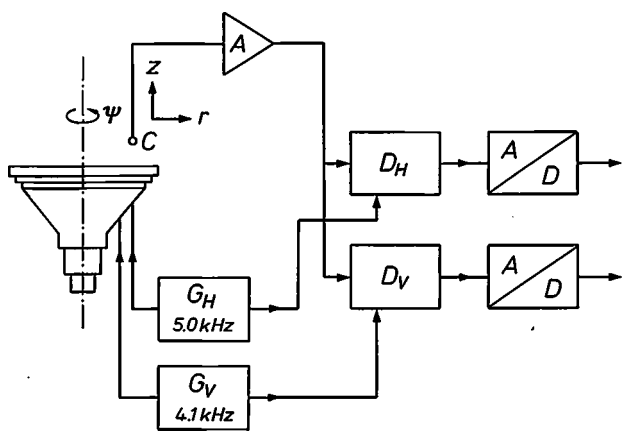


Fig. 2. Block diagram of the measuring equipment. The deflection unit can be rotated (the ψ-displacement), and the probe C can be moved both horizontally (the r-direction) and vertically (the z-direction). A sinusoidal signal of frequency 5.0 kHz is applied to the horizontal deflection coil, and a sinusoidal signal at 4.1 kHz is applied to the vertical deflection coil. After amplification, the signal containing components at these two frequencies is synchronously detected (D_H, D_V) and the two resulting signals are converted into digital signals for processing by a P855 minicomputer.

The complete measurement consists of a number of measurements for successive circles. For each circle the probe is set in the appropriate r,z-position and the turntable is set to the starting position, i.e. $\psi = 0°$ or $\psi = 360°$, depending on the direction of rotation. (After the measurements for each circle the direction of rotation is reversed so that the leads to the deflection unit do not get wound up.) A maximum of 144 points can be measured for each circle (360°/2.5°), but to speed up the measuring process the number of points actually measured is kept as small as possible. The number measured depends on the rate at which the fields vary in space and is thus correlated with the position of the circles: near the windings of the deflection unit the number is often 72 per measured circle, but in the direction of the screen it slowly decreases to 8 points per circle. In a complete measurement values are determined on about 60 circles with a total of some 2400 data points. The measurement takes about an hour and a half.

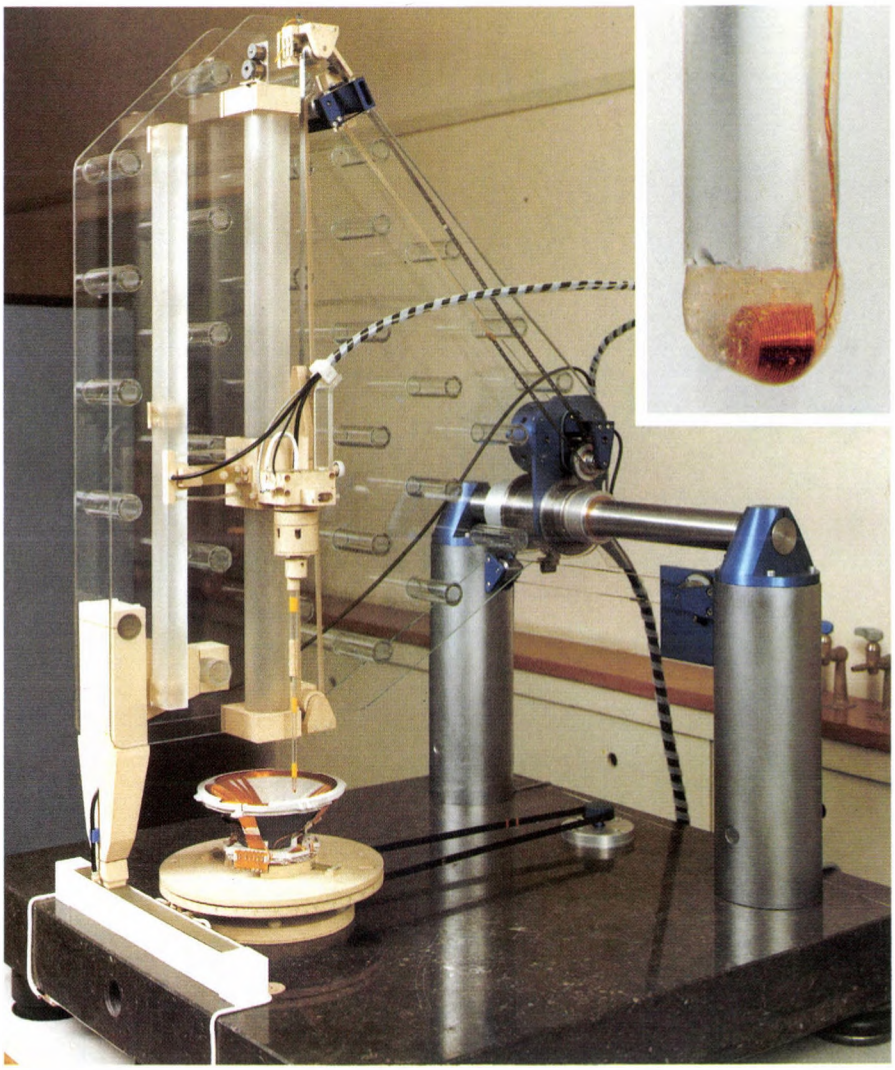


Fig. 3. Photograph of the equipment. The deflection unit is attached to a turntable, which provides the ψ-displacement. The probe is attached to a glass rod and can be moved vertically (in the z-direction) along a glass column. The column is attached to a double glass plate, giving a very stiff construction. The complete assembly can be moved horizontally along a metal shaft (in the r-direction); see fig. 4. The front of the assembly (on the left in the photograph) is supported on air bearings. There are no metal parts within 50 cm of the deflection coil. *Inset:* the probe. The length of the coil is 1.5 mm, the diameter 2 mm. These dimensions were chosen to ensure that the induced alternating voltage is determined as far as possible by the magnetic field-strength exactly at the centre of the probe.

After the measurements for each circle, the multipole components of the deflection field are determined from the results of the measurement by the P855 minicomputer. The components are entered on

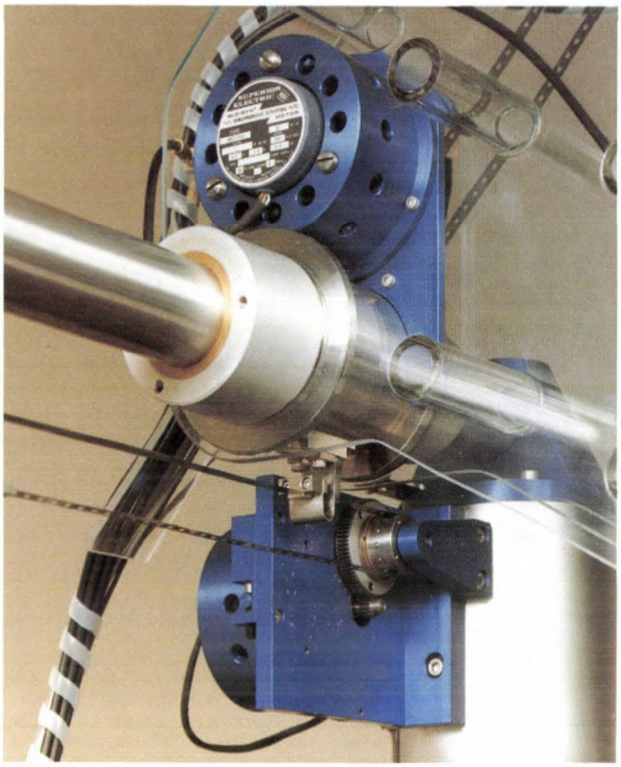


Fig. 4. Part of the equipment. The air bearing for the r-displacements can clearly be seen. Below it is the stepping motor which produces these movements. Above it can be seen the stepping motor for the z-displacements.

punched tape to permit further processing, such as a calculation of the field in the deflection region, on a larger computer.

The mechanical part

Fig. 3 shows the mechanical part of the equipment. The entire assembly is mounted on a granite block supported by air springs. To minimize interference with the magnetic fields to be measured, no metal is used in the construction within a distance of 50 cm from the deflection unit. The structural materials are mainly glass, ceramic and plastic.

The deflection unit is fixed to a turntable, which provides the displacements in the ψ-direction. The probe, shown enlarged in the inset of fig. 3, is located at the end of a glass rod, which can be displaced vertically along a glass column. This column is rigidly attached by two glass plates to the sliding part of an air bearing, so that the complete assembly can be displaced along a horizontal metal shaft. The front of the assembly rests on a sliding support, with an air bearing, that can move along a flat glass bar attached to the granite block. The assembly is moved horizontally by a stepping motor, visible in *fig. 4* below the first air bearing. The two glass plates are cemented together by a number of glass tubes, so that the assembly is relatively light, yet has the required stiffness.

Above the air bearing shown in fig. 4 there is a second stepping motor, which provides the vertical movement of the probe holder. Another air bearing allows the probe holder to be moved along the glass column. The probe holder is also supported against a flat vertical glass bar to prevent it from rotating. A safety switch ensures that the probe coil is disconnected if a programming error causes it to run against an obstacle.

The turntable

The construction of the turntable is shown in *fig. 5*. The rotating part is shaded grey. The accuracy of the turntable is better than 1 minute of arc; this is made possible by two intermeshing gear rings T. The teeth of the gear rings are cut extremely accurately and have a pitch of 2.5°. Since the upper gear ring is connected to the fixed part of the turntable assembly via a diaphragm that is rigid in the radial direction, the ring cannot move in the horizontal plane, but it does move in the vertical direction. To rotate the turntable R the upper gear ring is raised from the lower one by means of a lifting pin L. A third stepping motor then rotates the turntable R, including the lower gear ring, through an angle of 2.5° or a multiple of this, depending on the programming. Finally the pin L brings the gear rings into mesh again, thus accurately determining the rotation. Microswitches M give information about the

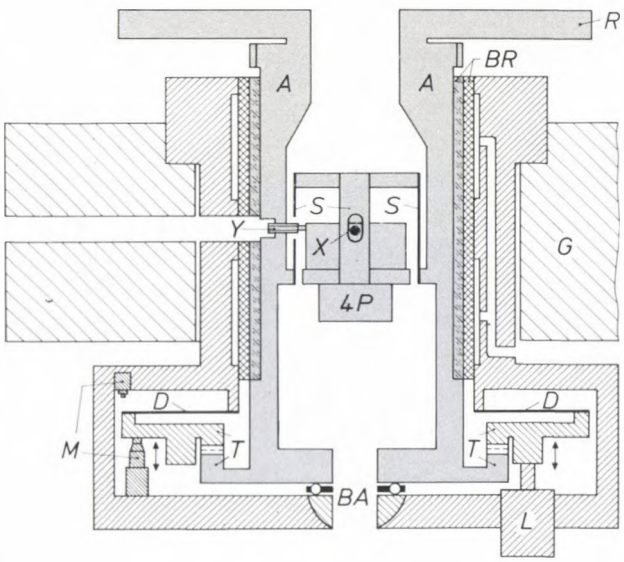


Fig. 5. Vertical cross-section through the turntable. The rotating part is shown grey. G granite block. R upper part of the turntable. BR radial air bearing, consisting of a glass inner cylinder connected to the central part A, and an outer cylinder of graphite; compressed air is supplied to the graphite cylinder. BA axial three-point bearing. T gear rings; the upper ring can be raised from the lower one by a lifting pin L. D diaphragm of great radial stiffness. M microswitches for detecting the position of the upper gear ring. $4P$ fourpole coil. S leaf springs. X and Y setscrews for the fourpole coil.

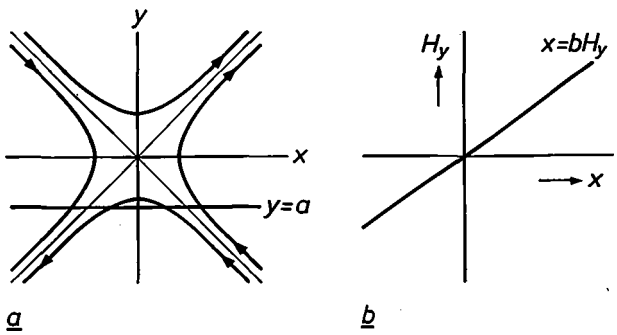


Fig. 6. *a*) Cross-section of the fourpole field in the x,y-plane through the point $z = 0$ perpendicular to the z-direction. The x-axis is in the direction of the r-displacement of the probe. When the probe is moved along a line $y = a$ and the field-strength H_y is measured in the y-direction, we have $x = bH_y$ (*b*). This relation is used for finding the correct position $r = 0$ (see text).

high dimensional accuracy, which is cemented to the part *A*, made of plastic. This prevents changes in the shape of the part *A* from causing unacceptable play in the bearing. The rotational accuracy of this air bearing is 2 μm.

The required axial accuracy (better than 10 μm) is achieved by means of a thrust bearing with three sapphire balls. This choice is related to the annular shape which is convenient for connecting the current leads to the deflection unit, etc. The lower part of the bearing is formed by a hemisphere that can adjust itself in relation to the baseplate.

Current leads are also required for the fourpole coil *4P* mounted at the centre of the turntable and suspended on two pairs of leaf springs. The function of

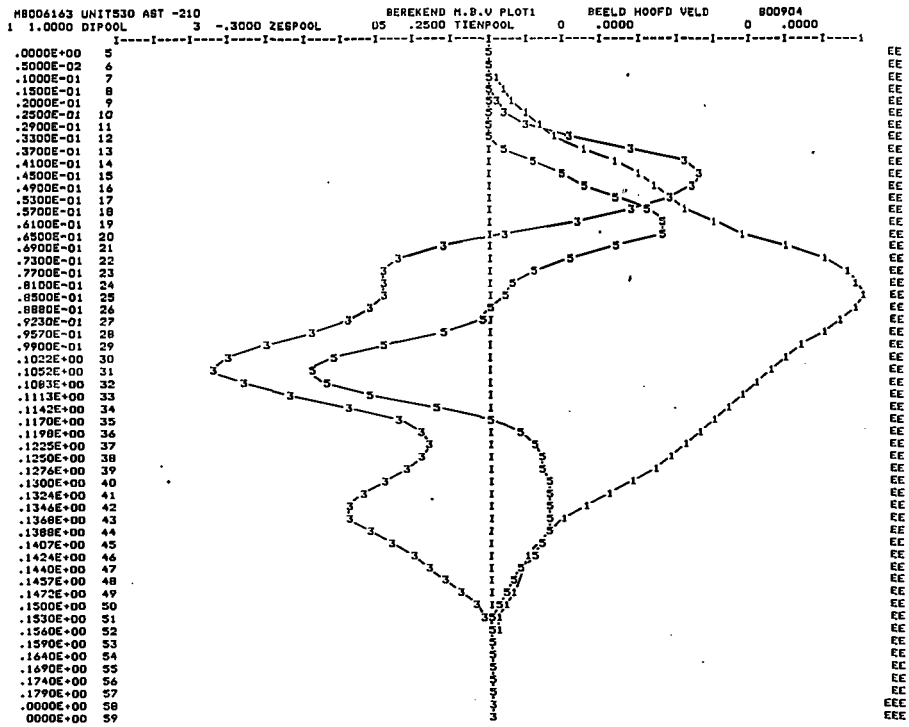


Fig. 7. Example of a graph plotted by a computer from a series of results of measurements relating to the vertical deflection field of a 30AX deflection unit. The vertical scale shows the numbers assigned to the measurement circles (in the second column). The curves show the amplitude variation of the multipole components of the field, with $n = 1$, $n = 3$, $n = 5$ [1]. In the design of deflection units like the one for the 30AX system even $n = 17$ is of some significance.

vertical position of the upper gear ring during these operations.

The deviations in the radial direction are also extremely small (< 10 μm). This is due to the use of a journal bearing *BR* of high quality. The outer, stationary part of the bearing consists of porous graphite; compressed air is supplied to the stationary part of the bearing through special channels. The inner part of the bearing is formed by a glass tube of the fourpole coil, which can be positioned by means of two setscrews *X* and *Y*, will be explained presently.

The three stepping motors are actuated from the P855 minicomputer. For simplicity 'blind' positioning is used, which means that instead of the actual position of the probe being measured, the nominal position is taken, as expressed by the number of steps of the motor. A complete step in the r- or z-directions corresponds to 0.1 mm.

Blind positioning makes it necessary to establish the starting position of the probe accurately. In the z-directions this is done by providing the drive belt of the appropriate stepping motor with a cam that operates a microswitch at a particular z-position of the probe. For the position $r = 0$ the fourpole coil mentioned earlier is used, which rotates with the turntable.

The reference point $r = 0$

An important characteristic of a fourpole field is that the amplitude from the centre, where the field-strength is zero, varies linearly with the distance to the centre in all directions. By positioning the fourpole coil by means of the two setscrews in such a way that its centre coincides with the axis of the turntable, it is possible to find the position $r = 0$ of the probe simply, since this is the only position where the field is zero.

The centre is found by means of the probe itself. *Fig. 6a* gives the situation for $\psi = 0$. The fourpole coil is at an angle of 45° to the x-axis. (In this case orthogonal coordinates are used, with the x,y-plane parallel to the plane of the turntable and with the x-direction coincident with the r-direction of the probe.) When the probe is moved along an arbitrary line $y = a$ in the x-direction, x and the measured field-strength H_y in the y-direction are related by $x = bH_y$ (fig. 6*b*). The x-position is first determined for which $H_y = 0$. Then the turntable is rotated through 180°, so that the same fourpole field is obtained, although at a different place because the fourpole coil and the axis of the turntable do not coincide. The distance through which the probe has to be moved to measure $H_y = 0$ again is read off and amounts to twice the eccentricity of the fourpole coil in the x-direction. The fourpole coil can now be aligned at the correct x-position. After the turntable has been rotated through 90°, this procedure is repeated to find the correct y-position.

The position $r = 0$ for the probe can now readily be found again at any time, because it is only in this position that we have $H_x = H_y = 0$ at any given ψ-position of the turntable. The setting accuracy with this method is better than 15 μm.

In the last few years this measuring equipment has proved a valuable instrument in bringing the 30AX concept to fruition. A result of one of the measurements that have made this possible can be seen in *fig. 7*.

Summary. In the design phase of a particular type of television deflection unit, such as the 30AX system, it is necessary to be able to find out how well the deflection fields approximate to the calculated 'ideal' deflection fields. Measurement of these fields will also provide an understanding of the coil parameters responsible for particular deflection errors. The article describes an instrument that can measure these fields with the accuracy required. The equipment is controlled by a P855 minicomputer, which measures the field at some 2400 points around the edge of the deflection region. The measurement takes about an hour and a half; the computer then determines the multipole components of the field from the measured data.

Scientific publications

These publications are contributed by staff of laboratories and plants which form part of or cooperate with enterprises of the Philips group of companies, particularly by staff of the following research laboratories:

Philips Research Laboratories, Eindhoven, The Netherlands	*E*
Philips Research Laboratories, Redhill, Surrey RH1 5HA, England	*R*
Laboratoires d'Electronique et de Physique Appliquée, 3 avenue Descartes, 94450 Limeil-Brévannes, France	*L*
Philips GmbH Forschungslaboratorium Aachen, Weißhausstraße, 51 Aachen, Germany	*A*
Philips GmbH Forschungslaboratorium Hamburg, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany	*H*
Philips Research Laboratory Brussels, 2 avenue Van Becelaere, 1170 Brussels (Boitsfort), Belgium	*B*
Philips Laboratories, N.A.P.C., 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A.	*N*

P. Ashburn & C. J. Bull: Observations of dislocations and junction irregularities in bipolar transistors using the E.B.I.C. mode of the scanning electron microscope.
Solid-State Electronics **22**, 105-110, 1979 (No. 1). *R*

G. Bartels & G. Passig: A new substrate holder for liquid phase epitaxy.
J. Crystal Growth **44**, 363-364, 1978 (No. 3). *H*

V. Belevitch: A class of nonrational impedances.
Int. J. Circuit Theory and Appl. **6**, 315-319, 1978 (No. 4). *B*

M. Boulou, G. Jacob & D. Bois: Cathodoluminescence study of Zn doped GaN.
Rev. Phys. appl. **13**, 555-563, 1978 (No. 11). *L*

A. Broese van Groenou: Pressing of ceramic powders: a review of recent work.
Powder Metall. Int. **10**, 206-211, 1978 (No. 4). *E*

S. D. Brotherton & A. Gill: Determination of surface- and bulk-generation currents in low-leakage silicon MOS structures.
Appl. Phys. Letters **33**, 890-892, 1978 (No. 10). *R*

R. Bruno, W. Hermann, H. Hörster, R. Kersten & K. Klinkenberg: The Philips experimental house: a system's performance study.
CCMS Rep. No. 85, pp. 249-263, 1978. *A*

R. Bruno & H. Hörster: What and where? Solar active systems or energy conservation in buildings.
CCMS Rep. No. 85, pp. 1-37, 1978. *A*

K. H. J. Buschow: Magnetic coupling of rare earth moments in amorphous alloys.
Solid State Comm. **27**, 275-278, 1978 (No. 3). *E*

M. M. Choy, W. K. Zwicker & S. R. Chinn (M.I.T., Lexington, Mass.): Emission cross section and flash-lamp-excited NdP_5O_{14} laser at 1.32 μm.
Proc. Int. Conf. on Lasers, Orlando 1978, pp. 163-167; 1979. *N*

M. Delfino, G. M. Loiacono, W. N. Osborne & G. Kostecky: Solution growth of L(+) hydrazonium tartrate.
J. Crystal Growth **46**, 241-244, 1979 (No. 2). *N*

P. Delsarte: Nombres de Bell et polynômes de Charlier.
C.R. Acad. Sci. Paris A **287**, 271-273, 1978 (No. 5). *B*

A. M. van Diepen & F. K. Lotgering: Mössbauer spectra of Fe^{3+} and Fe^{2+} ions in hexagonal ferrite with W-structure.
Solid State Comm. **27**, 255-258, 1978 (No. 3). *E*

J. A. W. van der Does de Bye, J. L. Sommerdijk, J. Hornstra, A. Bril & A. L. N. Stevels: A noncontinuum model for energy transfer in europium-activated oxytungstates.
J. Luminescence **18/19**, 285-288, 1979 (Part I). *E*

H. Fokker & J. A. M. van Eekelen: The description of the Stirling cycle in a vector diagram.
Proc. 13th Intersociety Energy Conversion Engng. Conf., San Diego 1978, Vol. III, pp. 1739-1745. *E*

H. Fokker & J. A. M. van Eekelen: Typical phenomena of the Stirling cycle as encountered in a numerical approach.
Proc. 13th Intersociety Energy Conversion Engng. Conf., San Diego 1978, Vol. III, pp. 1746-1752. *E*

H. M. Gibbs (Bell Laboratories, Murray Hill, N.J.) **& B. Bölger:** Coherent optical pulse propagation in thick resonant absorbers.
Coherence and quantum optics IV, ed. L. Mandel & E. Wolf, pp. 759-765; Plenum Press, New York 1978. *E*

H.-J. Hagemann: Reduction of hysteresis losses of ferroelectric $BaTiO_3$.
Ferroelectrics **22**, 743, 1978 (No. 1/2). *A*

M. L. Hermans & G. A. A. Asselman: A Stirling engine heat pump system.
Proc. 13th Intersociety Energy Conversion Engng. Conf., San Diego 1978, Vol. III, pp. 1830-1833. *E*

R. E. Horstman, J. Wolter & M. C. H. M. Wouters: Propagation of high-frequency phonons in thin SiO films at 60 mK.
J. Physique Lettres **39**, L 469-470, 1978 (No. 23). *E*

F. Hottier, R. Cadoret & J. B. Theeten: Caractérisation in situ par ellipsométrie d'un dépôt de Si polycristallin.
IVème Colloque de physique et chimie des surfaces solides, Antibes-Juan-les-Pins 1978 (No. spécial Le Vide), pp. 111-118. *L*

J. L. W. Kessels & A. J. Martin: Two implementations of the conditional critical region using a split binary semaphore.
Inform. Process. Letters **8**, 67-71, 1979 (No. 2). *E*

A. J. R. de Kock: Introduction of defects into silicon during growth and processing.
Physics of semiconductors, 1978 (14th Int. Conf., Edinburgh; Inst. Phys. Conf. Ser. No. 43), pp. 103-111; 1979. *E*

H. J. Leamy & A. G. Dirks: Microstructure and magnetism in amorphous rare-earth–transition-metal thin films: I. Microstructure, II. Magnetic anisotropy.
J. appl. Phys. **49**, 3430-3438, 1978 (No. 6), & **50**, 2871-2882, 1979 (No. 4). *E*

R. Linde, H. Weiss & W. J. Dallas: Single-emulsion-layer holographic inverse-filters for deconvolving coded-aperture images.
Proc. 1978 Int. Optical Computing Conf., London, pp. 135-138; 1979. *H*

F. P. Mattar*, M. C. Newstein*, P. E. Serafim*, H. M. Gibbs*, B. Bölger, G. Forster & P. E. Toschek**** (* U.S.A.; ** University of Heidelberg): Strong departures from uniform plane wave pulse propagation as a result of coherent transverse effects.
Coherence and quantum optics IV, ed. L. Mandel & E. Wolf, pp. 143-164; Plenum Press, New York 1978. *E*

W. C. P. M. Meerman & A. C. Knaapen: A high-pressure triaxial testing cell.
Powder Technol. **22**, 271-278, 1979 (No. 2). *E*

R. J. Meijer & A. P. J. Michels: Conceptual design of a variable displacement Stirling engine for automotive propulsion.
Proc. 13th Intersociety Energy Conversion Engng. Conf., San Diego 1978, Vol. III, pp. 1834-1840. *E*

A. van Oostrom & L. Augustus: Activation and early life of a pressed barium scandate cathode.
Appl. Surface Sci. **2**, 173-186, 1979 (No. 2). *E*

J. W. Orton & M. J. Powell: The Hall effect in semiconducting powders.
Phil. Mag. B **38**, 491-501, 1978 (No. 5). *R*

J. G. J. Peelen, B. V. Rejda* & K. de Groot* (* Free University, Amsterdam): Preparation and properties of sintered hydroxylapatite.
Ceramurgia Int. **4**, 71-74, 1978 (No. 2). *E*

J.-R. Périlhou: L'imagerie médicale par ultrasons et ses limitations.
J. fr. Biophys. Méd. nucl. **2**, 205-208, 1978 (No. 4). *L*

G. Prast & A. K. de Jonge: A free-piston Stirling engine for small solar power plants.
Proc. 13th Intersociety Energy Conversion Engng. Conf., San Diego 1978, Vol. III, pp. 1826-1829. *E*

J. Pribetich, M. Chive, E. Constant (all with Université de Lille I) **& A. Farrayre:** Design and performances of maximum-efficiency single- and double-drift-region GaAs IMPATT diodes in the 3-18-GHz frequency range.
J. appl. Phys. **49**, 5584-5594, 1978 (No. 11). *L*

W. J. J. Rey: Robust statistical methods.
Lecture Notes in Mathematics **690**, VI+128 pp., 1978. *B*

M. Schlindwein: Iterative three-dimensional reconstruction from twin-cone beam projections.
IEEE Trans. **NS-25**, 1135-1143, 1978 (No. 5). *H*

A. P. Severijns & F. A. Staas: An improved double circulation dilution refrigerator.
Progrès du froid aux très basses températures (Centenaire de la première liquéfaction de l'air), pp. 175-180; Inst. Int. du Froid, Paris 1978. *E*

F. A. Staas, W. van Haeringen & A. P. Severijns: The cooling power of ^{3}He-^{4}He dilution refrigerators.
Progrès du froid aux très basses températures (Centenaire de la première liquéfaction de l'air), pp. 169-174; Inst. Int. du Froid, Paris 1978. *E*

J. P. Stagg & M. R. Boudry: The neutralization of Na^+ ions in HCl grown SiO_2.
Rev. Phys. appl. **13**, 841-843, 1978 (No. 12). *R*

A. L. N. Stevels & J. A. W. van der Does de Bye: Models for energy transfer and their application to aluminate phosphors.
J. Luminescence **18/19**, 809-815, 1979 (Part II). *E*

Y. Tamminga, G. E. J. Eggermont, W. K. Hofker, D. Hoonhout*, R. Garrett* & F. W. Saris* (* FOM-Institute for Atomic and Molecular Physics, Amsterdam): Differences between ruby and Nd:YAG laser annealing of ion implanted silicon.
Physics Letters **69A**, 436-438, 1979 (No. 6). *E*

Q. H. F. Vrehen, H. M. J. Hikspoors & H. M. Gibbs (Bell Laboratories, Murray Hill, N.J.): Experiments on superfluorescence in cesium.
Coherence and quantum optics IV, ed. L. Mandel & E. Wolf, pp. 543-553; Plenum Press, New York 1978. *E*

J. H. Waszink & L. G. M. de Greef: Determination of temperature and plasma composition in a high-pressure CsI-Hg-Ar discharge.
J. appl. Phys. **49**, 5150-5154, 1978 (No. 10). *E*

F. W. Willmott, I. Mackenzie & R. J. Dolphin: Microcomputer-controlled column switching system for high-performance liquid chromatography.
J. Chromatography **167**, 31-39, 1978. *R*

PHILIPS TECHNICAL REVIEW

VOLUME 39, 1980, No: 11

Thirty-five years of cooperation with the Institute for Nuclear Physics Research (IKO)

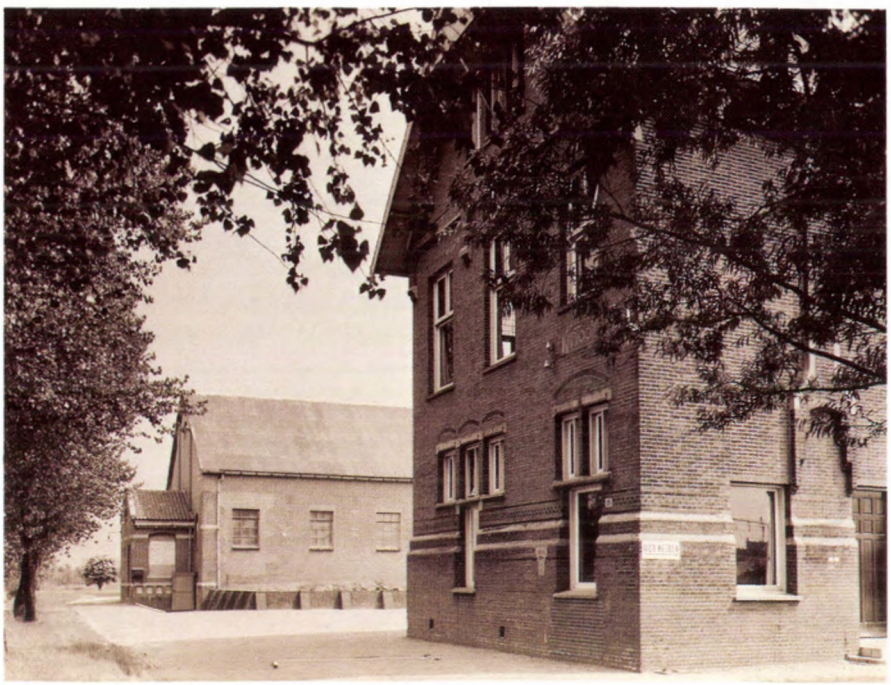
This issue of Philips Technical Review is a collection of articles that record the long cooperation between Philips and the Institute for Nuclear Physics Research in Amsterdam. (Its Dutch name is Instituut voor Kernphysisch Onderzoek, usually abbreviated to IKO.) Since the foundation of IKO in 1946 there have always been a number of Philips scientists working there, but because the interests of industry and academic nuclear research have gradually diverged through the years, the cooperative activity has now come to an end. IKO — which is now going to concentrate on nuclear physics at higher energies — became part of the National Institute of Nuclear and High-Energy Physics in January of this year, and the Philips group — now studying ion implantation — has moved to Eindhoven. The conclusion of this historic period provided a direct incentive for the publication of this issue, which gives an account of the more important results of the combined effort.

The introductory article looks back at the circumstances and considerations that led to the foundation of IKO. The articles that follow illustrate some of the important work carried out during the cooperative period. The series starts with the history and the development of the Philips synchrocyclotron, the first in Europe. This instrument was the basis for the birth of IKO and the start of a fruitful period of cooperation. A few scenes during the construction of this synchrocyclotron are shown in our 'snapshot album' near the centre of this issue. At first the synchrocyclotron was mainly used for producing radioisotopes; we have included a short article on this work. Next we come to Geiger-Müller counters and semiconductor detectors, instruments for observing radiation. The simultaneous use of a large number of semiconductor detectors in a single array for studying nuclear reactions is described in an article about IKO's 'BOL' system. Nuclear electronics for processing detector signals is the next subject. Mechanical engineering technology has always had an important part to play, and this is also the subject of an article. Finally, we look at some of the most recent work: ion-beam technology in the Philips group and IKO's new accelerator.

All of these articles have been compiled and edited by Dr J. E. J. Oberski of IKO. The hard work and enthusiasm he brought to this project is greatly appreciated by Philips Technical Review.

Although a few of the subjects discussed in the articles relate to a specific interest of either IKO or Philips, the emphasis in most of them is on a common interest. Prof. Casimir's 'science-technology spiral' can clearly be recognized: developments in scientific research, such as those at IKO, depend strongly on technological developments and also stimulate them in turn. By no means all of the research subjects have led to industrial production, but the knowledge obtained because difficult technical and scientific problems had to be solved meant that the cooperation was in many respects productive for Philips as well. Both Philips and IKO look back with satisfaction to the years of cooperation.

Philips tech. Rev. **39**, 286-289, 1980, No. 11



The old gasworks on the Ooster Ringdijk in Amsterdam, the building where IKO started in 1946.

The Institute for Nuclear Physics Research 'has finished its work'

J. M. Waalwijk and N. Wiedenhof

On the 29th of June 1946 the 'Instituut voor Kernphysisch Onderzoek' (IKO, the Institute for Nuclear Physics Research) was brought into being by three partners: N.V. Philips' Gloeilampenfabrieken, the Stichting Fundamenteel Onderzoek der Materie (FOM, the Organization for Fundamental Research on Matter) and the City of Amsterdam. Each partner made its contribution: Philips provided a synchrocyclotron — one of the first in the world and the first in Europe — the City of Amsterdam provided a freehold site with the buildings of an old gasworks on the Ooster Ringdijk, in Amsterdam, and FOM provided the money for a number of staff. The objectives of the new Foundation IKO were 'the advancement of fundamental and applied scientific research into nuclear physics and related subjects in the Netherlands, in the general interest and in the interests of higher education'.

Now, thirty-five years later, the cooperative activity has come to an end. Why? Did something go wrong? Or is it just natural development? The authors of this article have delved into old documents and have talked with Prof. H. B. G. Casimir — who assisted at IKO's birth — and with Dr E. F. de Haan, who as Casimir's successor on IKO's Board of Governors, was involved in IKO's final metamorphosis: the transition, together with Dutch high-energy physicists, into a National Institute for Nuclear and High-Energy Physics (NIKHEF, for *Nationaal Instituut voor Kernfysica en Hoge Energie Fysica*), with the Philips group leaving IKO and going to the Philips Research Laboratories in Eindhoven.

Drs J. M. Waalwijk and Dr N. Wiedenhof are with the Philips Press Office, Eindhoven.

The heroic decade

The foundation of IKO depended on a great deal of history. Its birth was associated not with a single discovery, but a gradual development of a new discipline: nuclear physics. For Philips Research this new discipline had already begun to take shape in the thirties. 'A heroic decade for nuclear physics', Casimir had said of those ten years. In the previous decades scientific giants such as Antoine Becquerel, Marie and Pierre Curie, Ernest Rutherford and Frederick Soddy had mapped out the new field of research, and in the thirties there was an enormous expansion with a vast range of discoveries. James Chadwick discovered the neutron in 1932; in the same year Carl Anderson found the positron and H. C. Urey found the deuteron. In 1934 Enrico Fermi discovered that uranium bombarded with neutrons yielded radioactive elements of a higher atomic number: the transuranian elements had been discovered. Fermi took out a patent in 1935 for the use of slow neutrons in the preparation of radioactive materials. As early as 1936 — somewhat ahead of events — Philips became a licensee. The trail blazed by Fermi was taken further by Otto Hahn, Lise Meitner and Fritz Strassmann, and it led to the discovery of nuclear fission. It became clear to many that a whole new field of physics lay open, with even more important fundamental discoveries and — probably — with practical applications.

No wonder that the then Director of Philips Research Laboratories, Prof. G. Holst, showed interest. De Haan: 'Philips Research has always taken new disciplines into its programme if they seem likely to be of interest to our industry. The time at which to invest is important. Scientifically promising ideas supported too soon could turn out to be the hobbies of a small group of scientists.' But the interest in nuclear physics was given solid backing, because Philips were starting to strengthen their activities in 'professional' equipment at the time. They had already met with some success in carrier telephony, radio transmitters and X-ray equipment. In addition, X-ray therapy could perhaps be complemented by radiation therapy with radioactive materials, and then Philips could have something significant in this field both in instrumentation and — through Philips-Roxane, as it then was — in the production of radioactive material. The licensing agreements for the Fermi patent demonstrated Philips' early interest in these activities, and with hindsight it seems an obvious move. Casimir: 'In fact it wasn't so very obvious. It was Holst who should be given the credit for spotting things before they became obvious. But the agreement certainly made sense.'

High voltages at Philips

The direct involvement of Philips in the new area of research was also apparent from A. Bouwers's high-voltage studies, made on a special budget set up by Gerard Philips and continued by Anton Philips. At the time James Cockcroft — one of Rutherford's group at the Cavendish Laboratory in Cambridge — and Bouwers at Philips Research Laboratories in Eindhoven were working on a high-voltage generator operating on the cascade principle, which had been described by H. Greinacher in 1921. In the cascade generator an alternating voltage of a certain amplitude is induced in the secondary winding of a high-voltage transformer. A direct voltage at twice the amplitude can then be obtained by using a capacitor and a rectifier. This is repeated and in this way higher amplitudes can be obtained, with the voltage increasing in steps; hence the name.

Cockcroft originally used wet batteries to supply the heater current for the cathodes of his rectifiers. The difficulty here was that the batteries had to be continually checked. The arrangement was later improved by using small motor-driven generators. (These were very noisy, however.) Cockcroft and E. T. S. Walton demonstrated the value of the high-voltage installation in nuclear research by producing nuclear reactions with it.

Bouwers and A. Kuntke improved the cascade generator at Philips Research Laboratories, mainly by using radio-frequency heating for the rectifier cathodes. Cambridge were impressed by the improved experimental model and purchased a Philips generator in the mid-thirties. This generator had a particularly successful career at Cambridge and was eventually donated to the University of Johannesburg where it continued to give many years of useful service.

Philips continued to supply various cascade generators even after World War II.

The cyclotron appears

In the meantime C. J. Bakker and F. A. Heyn were making a further reconnaissance into nuclear physics at Philips Research Laboratories. In 1939 they were joined by the chemist A. H. W. Aten, Jr. They made use of Bouwers's high-voltage installation and the neutron generator built at the Laboratories to study various nuclear processes, and they prepared many radioactive materials, including the radioactive phosphor now widely used in radiochemistry.

And so the idea gradually arose of building an accelerator, a cyclotron, in order to have a much more powerful machine available for nuclear research, a machine that could generate particles of higher

energy than had so far been possible. Hadn't the American physicist E. O. Lawrence of the University of California at Berkeley, the father of the cyclotron, achieved great success with the cyclotron he had constructed back in 1929, and hadn't he even been given the Nobel Prize for it in 1939? Bakker and Heyn set to work and made models, calculations and drawings. In the meantime, World War II broke out in all its fury, before they were ready to order parts from the mechanical-engineering departments. Various activities had to take place without the knowledge of the occupying Germans, since they had already turned down applications for material because a cyclotron was *'nicht kriegswichtig'* (of no military value).

The setbacks of those days had their advantages, for when the war was over, Bakker and Heyn's plans formed the basis for a more advanced design, the synchrocyclotron, which could take into account the relativistic increase in mass of particles at very high velocities. Immediately after World War II accelerators were the most advanced scientific instruments. In De Haan's words: 'Nuclear physicists were in a state of euphoria. Everyone wanted to work with a cyclotron. There was a commercial market. And this was the apparatus that was the start of IKO.'

Looking for partners

Casimir: 'I had joined Philips in 1942 and Prof. Holst had asked me if I would just look into a magnet for a cyclotron. So I pondered what to do, but very soon decided to leave the subject to Bakker and Heyn, since they knew far more about these things than I did. And after the war — I had just become a Director of Philips Research Laboratories — Philips interest in this had diminished somewhat... It seemed sensible in any case to seek partners for such a project. I did not think setting up such an extensive nuclear-physics activity for Philips could be justified, and I certainly did not have nuclear energy in mind. Philips didn't go in for 'high energy' — as did Siemens and General Electric. We didn't build power stations, water turbines or large electric motors. It wasn't our line of country at all.

What's more, Philips wanted to be as independent of other suppliers as possible. In no way did we want to be dependent on supplies of uranium, with all the complications that this could bring. But I did see possibilities of setting up a joint project with a university. The time was ripe for this when Bakker became Professor of Physics at Amsterdam. Perhaps it was he who set the ball rolling for IKO, probably because he didn't want to leave the cyclotron that he and Heyn had put so much work into. In this way you could do highly specialized work without the expenses being too high. Philips would be able to take advantage in the product sector, not only by manufacturing cyclotrons — which would only have a limited market anyway — but by making various kinds of instrumentation as well. And that's what happened.'

IKO gets under way

And so IKO came into being as the provisional objective of a long preparatory phase. On the 19th of September 1949 the first Dutch (synchro)cyclotron, housed in a former gasworks in the Watergraafsmeer in Amsterdam, produced its first beam of charged particles. Reading the old accounts of the event shows what an important moment this was for the IKO people. The official opening took place on the 10th of November 1949, in the presence of the Ministers of Economic Affairs and of Education, Arts and Sciences, the Philips top management — P. F. S. Otten, G. Holst, H. B. G. Casimir, H. Rinia, H. Bienfait, the Burgomaster of Amsterdam, A. d'Ailly, and leading scientists from Switzerland, Belgium, the Netherlands, Denmark and Sweden.

Prof. Schermerhorn was also present as the representative of the first Dutch Government after the liberation, which had supported the plans for IKO.

Prof. Clay, who was then the Chairman of the IKO Board of Governors, closed his opening speech with these words — and they give a clear picture of the expectations — 'Ladies and gentlemen, science in the Netherlands has acquired an exceptionally important instrument. I know that many young physicists are eagerly looking forward to the chance of working here. Under the first-rate guidance of the gentlemen we have managed to bring to this laboratory, the work will most certainly be productive and will continue to uphold the high traditions of our physical science.'

Bakker, originator and worker from the very start, also addressed the guests. He finished by saying 'It is my firm conviction that a country that neglects the fundamental study of matter deprives itself of the basis for the development of its industries. Industry profits directly from the results of fundamental research that come into consideration for technological development.' High hopes indeed.

Further activities

In 1958 a new landmark was reached: the extraction of a beam of particles from the cyclotron. Previously, targets had been inserted *inside* the cyclotron casing so that the material could be bombarded by the particles

accelerated in the cyclotron. However, for studying various nuclear reactions a beam of fast particles *outside* the cyclotron is required. This called for a fair amount of experimental effort, but by 1958 success had been achieved.

In 1960 the machine was considerably modified and then transferred to IKO for the token price of one guilder.

In the meantime — from the commissioning until about 1958 — many radionuclides were manufactured for Philips-Duphar, including radioactive cobalt that was used in cobalt needles for cancer therapy, and radioactive iodine for thyroid investigations.

From 1950 to 1960 the Philips group at IKO applied their efforts to the instrumentation, as Casimir had foreseen. They designed Geiger-Müller counter tubes that were later produced in large numbers by Elcoma in Eindhoven. The birthplace of these activities was Amsterdam. After 1960 attention shifted to the semiconductor radiation detectors and these have also had a commercial sequel at Elcoma.

The parting of the ways

The market for cyclotrons was short-lived, as we now know. Evolution in nuclear-physics research was rapid, so that there was soon a need for accelerators for higher energies. And the IKO people were beginning to think about new projects. They opted for a linear accelerator for electrons (maximum energy 500 MeV). This was a move away from low-energy physics towards high-energy physics, and also a move away from the interests of their research partner. But Philips were also thinking about their future too. De Haan: 'We owe much to Dr W. K. Hofker here, who pointed out that ion implantation would be an interesting field for Philips. Surfaces and surface treatment are of great importance for us, and ion implantation was a promising method for directly modifying the structure.'

From that time — the late sixties and early seventies — the parting of the ways became more clearly marked. The interests of IKO and Philips in physics began to diverge. Ion implantation soon turned out to be of great practical significance for Philips, and in the early seventies it was decided to transfer the work to Philips Research Laboratories in Eindhoven. The actual move was postponed till 1980 for various internal reasons. Now IKO will become part of NIKHEF, and the partnership is formally at an end.

Casimir: 'It is a really interesting evolution, when we look at it. At first people in a number of university laboratories were working on high-voltage installations. Then Philips came along and Bouwers and Kuntke made advanced high-voltage generators. Next Philips produced a number of cyclotrons. There was a limited market for these, but certainly not a continuing one for quantity production. In the meantime there had been a considerable evolution in nuclear-physics research, and the research apparatus had become extremely complicated systems. Indeed, it had gone beyond the capacity of ordinary industry — think of those vast machines at CERN, the European research centre in Geneva. And this brought the industrial production of such machines to a fairly natural conclusion.'

Just as the time gradually became ripe for the start of a cooperative relationship, so the time came for wondering whether cooperation still made sense. In a maturing subject all sorts of branches appear, and it may be more sensible to let particular specializations go their own way. De Haan looks back with satisfaction on the 35 years of cooperation with IKO. 'We learned a lot from it. And as for the future, I don't think it impossible that we might be in touch with our old partners again at NIKHEF. I am thinking of things like software or research with fast electrons, which we are interested in. But that is for the future. Looking back, I can truly say that I wouldn't have missed the fun for anything.'

Philips tech. Rev. 39, 290-292, 1980, No. 11

The cyclotron

G. Luijckx

The success of the concept of using a high-voltage installation to accelerate particles was quickly followed by the idea of a cyclic accelerator. In this system particles pass through the same accelerating field several times, so that a high energy (e.g. 10 MeV) can be obtained with an alternating voltage of relatively low amplitude (e.g. 10 kV). In about 1930 E. O. Lawrence and M. S. Livingston constructed a proton accelerator based on this principle in Berkeley, California: the first cyclotron. They obtained an energy of 80 keV. A few years later Berkeley had a cyclotron delivering deuterons at 8 MeV, whereas the Cockroft-Walton high voltage accelerator in Cambridge only gave about 1 MeV. Shortly afterwards Berkeley even managed to reach 25 MeV.

Attempts to get above this energy, however, were hampered by the relativistic increase in the mass of the accelerated particles. A plan to go beyond this limit at Berkeley failed to materialize because of the outbreak of the Second World War. The intention was to use a higher accelerating voltage, which would have meant that the particles would need to make fewer revolutions to attain the same energy. After the war, however, observations by V. I. Veksler (U.S.S.R.) and E. M. MacMillan (U.S.A.) showed that the particles remain in phase with the accelerating voltage if either the acceleration frequency or the strength of the magnetic field in which the acceleration takes place varies slowly. Use of these methods to compensate for the effect of the relativistic increase in mass led to new types of accelerator — the synchrocyclotron and the synchrotron — which would eventually give energies in the GeV range. In 1947 Berkeley had a synchrocyclotron that delivered deuterons of 195 MeV and helium nuclei of 390 MeV; in 1957 this machine could accelerate protons up to an energy of 720 MeV.

Some aspects of the history of the Philips cyclotron — with which the names of C. J. Bakker and F. A. Heyn are particularly associated — are described in the introductory article [1] of this issue. Philips' original plan for the construction of a cyclotron was modified to take advantage of the latest information and the Philips synchrocyclotron [2] at IKO was completed in 1949. After extensive modifications the synchrocyclotron was donated to IKO in 1961. One of the modifications made was to replace the oil-cooled magnet coils by water-cooled aluminium coils. Various other changes were to be made later so that the equipment could meet the latest requirements for nuclear-physics research. It was finally shut down in 1977 so that all the Institute's manpower could be employed on the most recent accelerator development at IKO: the construction of a linear electron accelerator [3].

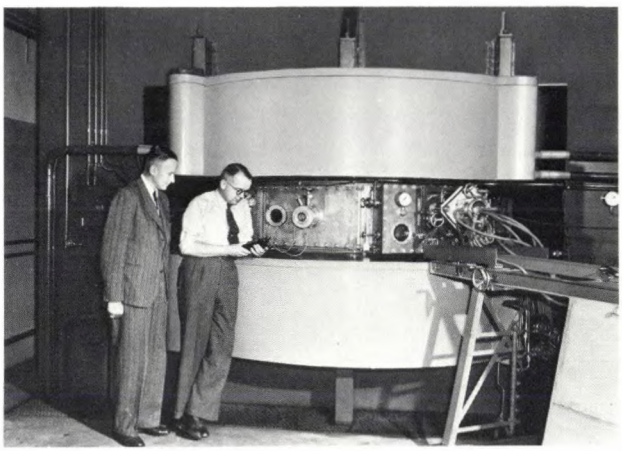
By that time more than thirty thousand targets had been irradiated and about twenty new radioisotopes had been found with the aid of the Philips synchrocyclotron. The research had produced hundreds of scientific publications and numerous doctorates. The synchrocyclotron had proved to be a valuable instrument.

Ir G. Luijckx is with IKO (Institute for Nuclear Physics Research), Amsterdam.

Operating principle of a cyclotron

A cyclotron (*fig. 1*) contains two hollow accelerating electrodes, the dees. An alternating voltage is applied between the dees, which are located between the poles of a magnet. The dees are enclosed in a vacuum chamber to prevent unwanted effects due to collisions between the accelerated particles and air. At the centre of the vacuum chamber, a gas — for example hydrogen gas if hydrogen nuclei (protons) are to be accelerated — is ionized in the 'ion source' [4]. The ions are accelerated in the electric field between the dees. Under the action of the Lorentz force they describe a circular orbit in the cavity formed by the dees. The electric field between the dees is reversed so rapidly that the next time the particles cross from one dee to the other the particles are subjected once again to an accelerating force. (For example, to accelerate protons in a magnetic field of 1.4 T requires an acceleration frequency of 20 MHz.) The higher velocity, however, corresponds to a circular orbit of a

greater radius until ultimately the particles reach the outer boundary of the magnetic field and no further acceleration is possible. This means that there is a maximum to the velocity — or energy — that can be achieved.



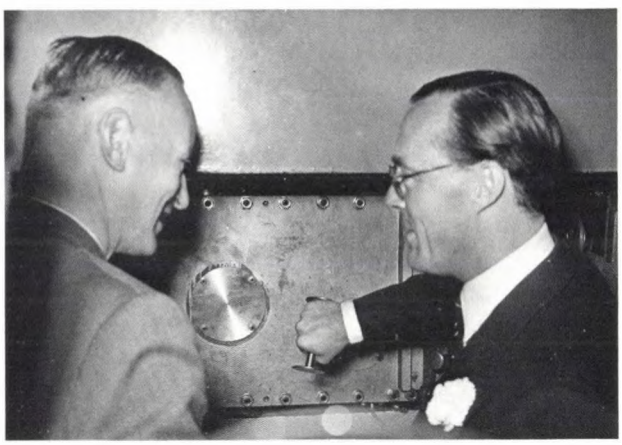


Fig. 1. *Above:* Prof. Bakker and Prof. Heyn — holding an instrument — with the brand-new synchrocyclotron. *Below:* HRH Prince Bernhard testing the power of the magnetic field of the synchrocyclotron during its official commissioning in 1949. Prof. Bakker is on the left.

Synchrocyclotron

When particles are accelerated to higher energies the relativistic increase in mass becomes important. This makes the angular velocity decrease gradually, until eventually the particles cross over between the dees when there is a decelerating field, and this interferes with the acceleration process. In a synchrocyclotron this effect is prevented by modulating the acceleration frequency so that it decreases synchronously with the orbital frequency. In this way it is possible to obtain a higher energy of the particles. The method has the disadvantage, however, that only one group of particles can be brought from zero to maximum energy in each modulation period. Particles are delivered in pulses, and the time interval between two successive pulses is determined by the modulation period.

Since 1950 it has been possible to get over this problem. By increasing the strength of the magnetic field appropriately as the radius increases it is possible in principle to continue accelerating at constant frequency in spite of the relativistic increase in mass. In such a magnetic field, however, the particle orbits are not stable in the transverse direction: any small transverse path deviation continues to increase. Stability can be restored by modulating the field-strength in the azimuthal direction and increasing the mean field-strength as the radius increases. This leads to AVF cyclotrons (AVF: Azimuthal Varying Field).

External beam

In the early days the Philips synchrocyclotron at IKO was used for making radioisotopes by introducing a target into the vacuum chamber (using a special vacuum lock) and irradiating it with particles. The energy of the particles was varied by selecting the orbital radius where the target was located.

Later there arose a keen interest in the study of nuclear reactions. These take place so quickly, however (10^{-22} s), that the decay products can only be observed during the irradiation. For this kind of nuclear experiment the accelerated particles had to be brought outside the synchrocyclotron ('beam extraction'). It was not easy to produce an external beam with a relatively small energy spread, however, because the particle orbits corresponding to different energy values are less than a millimetre apart. After an attempt at electrostatic deflection had failed, it was decided to use a 'regenerator', a highly suitable iron construction based on K. J. Le Couteur's machine in Liverpool. In this arrangement the magnetic field is perturbed at its outer edge in such a way that the particles begin to vibrate strongly in the radial direction. After the particles have passed through the regenerator several times the distance between two adjacent orbits is large enough to allow only the particles in the outermost orbit to leave the accelerator in a path determined by control magnets. Five to ten per cent of the internal particle current could be extracted in this way, a percentage quite normal for that time. (In later cyclotrons almost 100% has been achieved.)

[1] See J. M. Waalwijk and N. Wiedenhof, The Institute for Nuclear Physics Research 'has finished its work', this issue, page 286.
[2] F. A. Heyn and J. J. Burgerjon, The synchrocyclotron at Amsterdam, I, II, III, IV, Philips tech. Rev. **12**, 241-247, 247-256 and 349-364, 1950/51, and **14**, 263-279, 1952/53.
[3] See for example P. Bakker, Linear electron accelerators, this issue, page 325.
[4] See pages 267-270 of F. A. Heyn and J. J. Burgerjon, The synchrocyclotron at Amsterdam, IV, Philips tech. Rev. **14**, 263-279, 1952/53.

Stochastic extraction

Between 1965 and 1970 the BOL [5] nuclear-research system was developed at IKO. It was specially designed to observe two or more reaction products or 'coincidences' from a single nuclear reaction. To minimize the problems that might be caused by random coincidences — the simultaneous observation of two or more reaction products of different nuclear reactions — the particle current should have as few peaks as possible, since the number of random coincidences is proportional to the square of the current, whereas the number of real coincidences is proportional to the current. Because of the modulation, however, a particle current leaves a synchrocyclotron for only ten per cent of a modulation period: a more uniform distribution was therefore desirable. This modification was achieved by giving the accelerator an extra accelerating electrode, the 'cee', with its associated frequency generator and a high-speed high-voltage switch (pulser) to switch off the dee oscillator when the particles had almost reached the maximum energy. The acceleration was then taken over by the cee, with a modulated acceleration frequency, so that the particle orbits were spread out. The acceleration frequency of the cee then slowly decreased, enabling the particles to be guided out of the accelerator in succession. This stochastic extraction continued, even though the high voltage at the dees had already been switched on again and a new acceleration cycle started (*fig. 2*).

Shutdown

After the synchrocyclotron was transferred to IKO in 1961 a number of other improvements and modifications were made. The angular spread of the external beam was guided by magnets towards screened-off experimental areas and, finally, a newly developed radio-frequency system [6] was installed for generating the accelerating voltage.

From 1973 onwards the synchrocyclotron gradually came to be used less and less and on the 1st April 1977 the accelerator was finally shut down. After all the modifications the only component remaining from the synchrocyclotron of 1946 was the iron of the magnet.

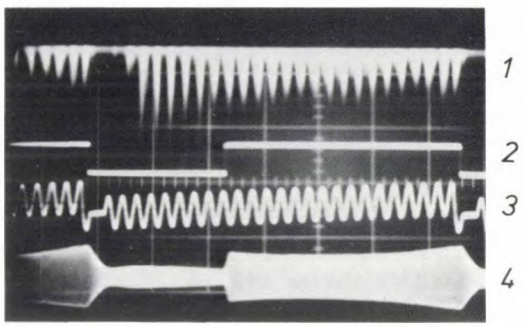


Fig. 2. Stochastic extraction. Oscillogram of the particle current *1* from the synchrocyclotron, the pulser voltage *2*, the frequency modulation *3* of the cee oscillator and the voltage *4* at the dee electrode. The time *t* is plotted on the horizontal axis; one scale division corresponds to 100 μs.

[5] See for example R. van Dantzig, BOL, this issue, page 302.

[6] The radio-frequency system consists of a resonant cavity, formed by the dee electrodes, a coaxial transmission line and a variable capacitor, the 'modulator'. This circuit can be considered as a resonant line with a shortening capacitor at each end. Altering the characteristic impedance of the transmission line changes the electrical length of the line and hence the resonant frequency. The circuit forms part of a power oscillator (Colpitts circuit). The r.f. system had a frequency range of 10 to 22 MHz, so that it was possible to accelerate four types of particle: protons, deuterons, helium-3 and helium-4 nuclei. Feedback at the various frequencies was obtained with the aid of large variable capacitors. Further details are given in W. van Genderen, Multi-particle rf-system for the IKO synchrocyclotron, Nucl. Instr. Meth. **54**, 288-292, 1967.

Casting one of the inner copper hemispheres for the BOL nuclear measurement project.

Radioisotopes

G. A. Brinkman

The synchrocyclotron at IKO came into operation in 1949. The nuclear particles accelerated in a cyclotron can be made to collide with atomic nuclei of a suitably selected material in a target placed in the cyclotron. Radioactive isotopes are then formed as a result of the collision.

By 1946 two Philips scientists, A. H. W. Aten (who was later to become the director of the Institute's radiochemical department) and F. A. Heyn, had already published two articles in Philips Technical Review on the use of a cyclotron in research: 'The use of isotopes as tracers' [1] and 'The technique of investigations with radioactive and stable isotopes' [2]. In 1948 Aten left Eindhoven for the Institute, and he was soon able to put his ideas into practice there. In 1951 he was followed by a team of seven, led by J. Halberstadt, from what was then Philips-Roxane. The group was given control over their own budget for laboratory space and chemicals; it was modest even for the time: about 10 000 guilders or $ 2600 at 1951 rates.

The chemical department of the Institute, who were doing fundamental scientific research, and Philips-Roxane, who were making isotopes for medical and industrial use, had a great many interests in common. These included finding the best nuclear reaction for making a particular isotope, designing suitable targets for the irradiation of particular materials, the purification of the irradiated material by means of fast separation techniques and the preparation of isotopes without a 'carrier' (added nonradioactive material with the same chemical properties as the radioactive material). This led to intensive cooperative use of the cyclotron, and this continued even after Philips-Roxane, who in the meantime had become Philips-Duphar, had purchased other premises in Amsterdam in 1959. In the period between 1950 and 1970 the cyclotron was run for more than 8500 hours, often at night, to provide irradiation for the production of radioisotopes for Philips-Duphar.

Another form of cooperation was the help given by Institute analysts in tasks such as cutting irradiated lead targets for the manufacture of bismuth-206 (20 guilders danger money was paid for this) and in processing isotopes such as iodine-131, caesium-137, iridium-192 and gold-198 that had been made in a nuclear reactor elsewhere. In 1958, for example, about 500 curies (1 curie = 3.7×10^{10} Bq), in the form of all kinds of isotopes, was despatched from the Institute by car, train, bicycle and aeroplane. A general survey of the activities of Philips-Roxane is given in the thesis by H. H. P. Moeken [3]. He describes the results of 45 nuclear reactions of deuterons with various materials as a function of the irradiation energy. Cost/benefit analyses are also given.

In 1954 the Netherlands Organization for the Advancement of Pure Research asked Philips-Roxane and the Institute to organize a joint course on isotopes. The course was to give people working with radioisotopes (doctors, hospital dispensers, students, salesmen) thorough and detailed information about radioisotopes and methods of using them. By 1980 about fifty two-week courses had been held; in recent years instruction courses have been given for examinations leading to the qualification 'Manager of a low-radiation-level radiological laboratory'. This cooperative scheme came to an end recently when the Philips-Duphar isotope activities were taken over by Byk-Mallinckrodt: the courses are now run entirely by the Institute.

Examples

A list will now be given of some of the more important isotopes that have been manufactured for medical and biological applications; some of them are still being manufactured today.

Sodium-24 (half-life $T_{1/2}$ = 15 hours)

Sodium-24 can be made from natural sodium in a nuclear reactor. When this isotope was required in the Netherlands, the irradiation was carried out in Kjeller, Norway and the isotopes were transferred from the Netherlands by air.

In about 1952 sodium-24 was required for osmosis research in Denmark, but without a 'carrier', i.e. pure radioactive sodium. When natural sodium is irradiated in a nuclear reactor the radioactive sodium

Dr G. A. Brinkman is with IKO (Institute for Nuclear Physics Research), Amsterdam.

This article arose partly out of discussions with W. B. Huising (formerly with Philips-Roxane and Philips-Duphar, now with Byk-Mallinckrodt) and J. Visser (IKO).

formed cannot be separated by chemical methods from the residual natural sodium — the major part of the end-product. The radioactive sodium was therefore made in Amsterdam by irradiating aluminium-27 with deuterons in the cyclotron, to form sodium-24 with the emission of a helium and a hydrogen nucleus. The radioactive sodium was separated from the remaining aluminium in a Dowex column (this took a great deal of time: at a flow rate of one drop per second the sodium was obtained from the column after five and a half hours, i.e. after 20000 drops). The isotope was then flown to Denmark; this meant that preparations of this isotope were travelling back and forth in aircraft flying between the Netherlands and Scandinavia.

Chlorine-34 ($T_{1/2}$ = 32 minutes), *chlorine-38* ($T_{1/2}$ = 37 minutes)

The isotopes were prepared by irradiation of sodium chloride, distillation of the chlorine into a sodium-iodide solution and boiling off the iodine. The solution had to be sterile, with the same osmotic pressure as human blood. The solution was used in the teaching hospital of the University of Amsterdam for research work on blood circulation.

Gallium-67 ($T_{1/2}$ = 3.2 days)

This isotope, made from zinc, was used for investigating bone tissue. Interest in the isotope gradually decreased, but has revived recently, this time for locating tumours other than in bones. The methods used earlier are again being used in its preparation.

Iodine-131 ($T_{1/2}$ = 8 days)

In about 1950 di-iodide fluorescein, tagged with iodine-131, was used as a detector and localizer in the investigation of tumours, with a dog as the experimental animal [4]. This investigation also involved the Physiological Laboratory of the University of Amsterdam.

Gold-198 ($T_{1/2}$ = 2.7 days)

Hundreds of curies of gold-198, prepared in a nuclear reactor, have been processed at the Institute into colloidal gold solutions stabilized with glucose and gelatin. The solution was used for locating liver and spleen abnormalities and for treating ascites in the abdominal cavity (as a result of malignant tumours), where the colloid adheres to the abdominal wall.

Bismuth-206 ($T_{1/2}$ = 6.2 days)

Formerly bismuth-206 was an isotope very much in demand, particularly for the therapy of chronic lymphatic leukaemia: in dissolved form ('liquid radium') it could be distributed through the lymphatic glands. When the cyclotron was built at Philips-Duphar, it was expected that almost half of the production time would be devoted to this isotope. Within two years, however, all interest had vanished.

Thallium-201 ($T_{1/2}$ = 3 days)

This isotope is much in use at the present time, particularly for heart studies. It is prepared by irradiating thallium with protons in a cyclotron and by separating the lead-201 formed ($T_{1/2}$ = 9.4 hours). The lead-201 decays into thallium-201 and this is then isolated again.

Today, Byk-Mallinckrodt's modern factory in Petten remains as an indirect inheritance — via Philips-Roxane and Philips-Duphar — of the cooperation between Philips and IKO in radioisotope work, an activity that started in about 1950 through the interest and efforts of a tiny group of people.

[1] A. H. W. Aten Jr. and F. A. Heyn, Philips tech. Rev. 8, 296, 1946.
[2] A. H. W. Aten Jr. and F. A. Heyn, Philips tech. Rev. 8, 330, 1946.
[3] H. H. P. Moeken, Thesis, University of Amsterdam, 1957.
[4] A. H. W. Aten Jr., J. Halberstadt and G. P. M. Horsten, Rev. méd. Liège 5, 428, 1950, and Acta physiol. pharmacol. neerl. 1, 331, 1950.

Geiger-Müller counters

W. K. Hofker

Once the synchrocyclotron at IKO had been completed, equipment was required for measuring radioactive radiation. This was necessary primarily for monitoring the level of any radiation that the staff might be exposed to, and secondly for the research activities, then mainly radiochemical. Philips was particularly interested in the development and production of test instruments such as gas-filled counter tubes, i.e. radiation detectors operating on the principle of ionization processes in gases.

Fig. 1. Some Geiger-Müller counter tubes developed by the Philips group at IKO. *1*, *2* and *3* counter tubes for α- and β-radiation; *4*, *5* and *6* counter tubes for γ- and X-radiation; *7* 4π liquid counter; *8* counter tube for α- and β-radiation, for use in guard tube *9*. Tube *4* is 400 mm long, 38 mm in diameter.

A Geiger-Müller counter tube contains an anode wire mounted coaxially inside a cylindrical cathode; an electrical potential is maintained between anode and cathode. The filling is a rare gas. An ionizing particle originating from, say, a radioactive specimen forms a track of ions and electrons in the gas filling, and these are accelerated in the electric field. Secondary ionization due to collisions of the primary ions and electrons with the gas atoms then sets up a gas discharge. The tube can count the *number* of particles captured, but it cannot measure the *energy* of the particles, because there is no relationship between the energy and the number of ions and electrons produced. (The energy can be determined, however, by using proportional counters.)

This detection method was employed as early as 1908 by Ernest Rutherford and Hans Geiger and has been the subject of much research. One of the important aspects studied was the method of quenching the discharge. At first this was done by reducing the voltage across the electrodes by using a high series resistance or by using an electronic switching circuit. Later, in 1935, A. Trost discovered that adding alcohol vapour to the rare gas quenched the discharge, even with a smaller series resistance: the alcohol molecules inhibit the escape of new particles from the cathode so that the discharge stops. However, during the discharge the alcohol is broken down, and this reduces the life of the counter. In 1937 Geiger and O. Haxel found that the life could be increased if a halogen was used as the quenching gas: the halogen fragments appear to recombine again.

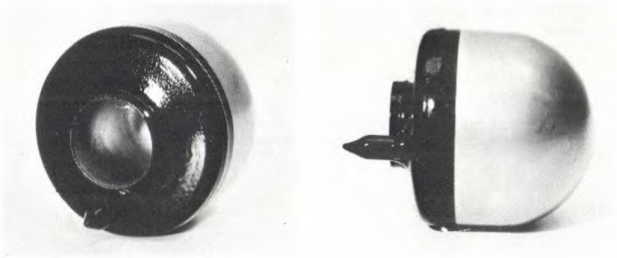
The research at Philips Research Laboratories was mainly on halogen counters and was carried out in

Dr Ir W. K. Hofker is with Philips Research Laboratories, Amsterdam Department.

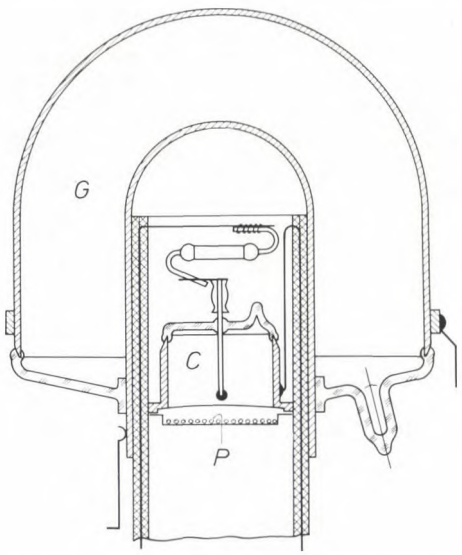
Eindhoven [1] and Amsterdam [2]. This research led to the production of a large number of different types of gas counter tubes (*fig. 1*).

Some results of the research

The aspects investigated by the Philips group in Amsterdam included the discharge mechanism, the counting efficiency and the measurement reproducibility of gas counters [2]. For a Philips 18506 counter, for example, they determined the behaviour of the quantity of charge released per pulse as a function of the counter voltage and the anode impedance. It was found that the charge was nearly proportional to the voltage and was only slightly affected by the impedance. Investigations were also made of the dead time — the time immediately following a detection during which the counter does not respond. A short dead time was achieved with a large counter voltage and a small parasitic capacitance of the anode. It was also found that with the small anode capacitance the measurement reproducibility was high, since the counting efficiency was not so dependent on the supply voltage.



a



b

Fig. 2. *a*) Guard counter tube. *b*) Diagrams showing guard tube *G* with its central counter tube *C* for β-radiation. *P* radioactive sample.

Besides the electronic circuit, the geometrical shape also had a considerable effect on the action of the counter. Earlier counters had always had a thin anode wire. The discharge in such a counter starts close to the anode where the field-strength is highest. The addition of a small quantity of halogen gas to the rare gas and the use of a thick anode wire was found to improve the counting efficiency, because the discharge could now start anywhere in the gas volume. The signal from the counter therefore reached its maximum value more rapidly.

This discovery led to new types of counter tube: the thick anode could now be made hollow. A radioactive source can be measured with a solid angle of 4π in the hollow cavity and radioactive liquids can be made to pass through it. It is even possible for one counter tube, the central counter tube, to be fitted into the hollow anode of another counter tube, the guard counter, and this arrangement (*fig. 2*) can be used to measure very weak beta radiation or soft X-radiation. If no special measures are taken, background radiation can interfere with the measurements: this background radiation consists of mesons (cosmic radiation) and gamma radiation originating from natural radioactivity in the environment, e.g. in building materials. The measurement equipment can be shielded from gamma radiation, e.g. with lead, but not from mesons. A meson that triggers a discharge in the central tube must necessarily have also triggered the guard counter tube; this effect can be used to reduce the contribution from such interference: an electronic anti-coincidence circuit selects the discharges in the central tube that do not correspond to simultaneous discharges in the guard counter tube [3]. In this way the observed intensity of the background radiation can be reduced by a factor of 20. The simple and sturdy construction also gives a high measurement reproducibility.

Hundreds of instruments of this kind are now in use all over the world for applications such as monitoring radioactive contamination of water, milk and other foodstuffs.

In high-energy and nuclear physics further developments resulted in 'wire chambers': gas-filled counters with hundreds of anode wires and a detection-sensitive area of many square metres, which are used to determine the point of entry of an ionizing particle. By using more 'wire chambers' it is also possible to measure the direction of travel of the particle. The large amount of data obtained per detection is processed by fast electronic circuits and computers.

[1] N. Warmoltz, Philips tech. Rev. **13**, 282, 1951/52.
[2] A review of research on gas-filled counters is given in K. van Duuren, A. J. M. Jaspers and J. Hermsen, Nucleonics **17**, No. 6, 86, 1959.
[3] K. van Duuren, Philips tech. Rev. **20**, 170, 1958/59.

Semiconductor detectors

W. K. Hofker

Between 1958 and 1968 we in the Philips group at IKO were working on the development of semiconductor detectors. This kind of detector is much more satisfactory for measuring various types of radiation than counters such as the Geiger-Müller counter. A Geiger-Müller counter can count the number of particles, but it cannot determine the energy of the particles [1]. This is because the charge it supplies to the electronic circuit is not in general proportional to the absorbed radiation energy. In the semiconductor detector this proportionality does exist, because the material is a solid. Furthermore, the absorption capacity of a solid is much greater than that of the counter gas. The volume of a solid-state detector can therefore be much smaller.

The first solid-state detectors used the scintillations caused by radiation in certain substances — for instance sodium iodide activated with a little thallium. The scintillations were converted into an electrical signal by a photomultiplier tube. Another type of detector, the conduction counter or 'crystal counter' [2] made its appearance when it became possible to make purer crystals and more sensitive amplifiers. In this type of detector the solid (e.g. semiconductor material) is provided with conducting contacts. The charge released in the solid by the radiation is measured directly.

The ability to make semiconductor detectors that had a diode characteristic (a *P-N* junction) led to a considerable improvement in the energy resolution. In addition, these semiconductor detectors are often faster than the ones mentioned above. If a sufficiently high voltage is applied to the semiconductor diode in the reverse direction, a region is created that has a high electric field-strength and hence few free charge carriers: the depletion layer. An ionizing particle that penetrates into this region forms a track of electron-hole pairs. Under the influence of the field the electrons and holes move in opposite directions towards the electrodes, and a current pulse is produced in the external circuit. The electric charge in the pulse is a measure of the energy that the particle has given up in the depletion region.

The semiconductor detector was introduced by K. G. McKay [3] as early as 1949, but not until 1956 did it attain practical importance in the form of the surface-barrier detector [4]. A detector of this kind consists of a wafer of silicon that has a high resistance and has a thin layer of metal applied to it (*fig. 1*). A Schottky barrier is formed between the silicon and the metal layer. When a reverse voltage is applied a detection-sensitive layer about 0.5 mm thick is formed; this is sufficient for measuring alpha radiation. However, the layer is too thin to detect gamma radiation or X-radiation.

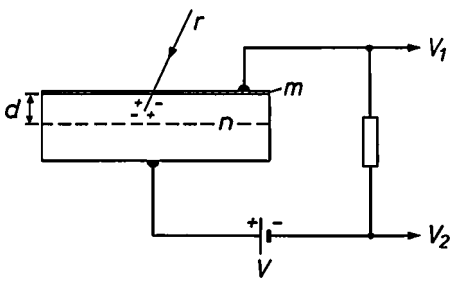


Fig. 1. Diagram of a surface-barrier detector with a voltage applied to it in the reverse direction. *m* metal electrode, usually a thin gold layer that has been deposited by evaporation. *n* *N*-type substrate. *d* thickness of the layer in which there is a strong electric field. The thickness of this layer is proportional to $\sqrt{\varrho V}$ where ϱ is the resistivity of the basic material and V is the voltage across the detector. The electrons and holes released by an ionizing particle *r* in the layer of thickness *d* produce a current pulse that is applied to an amplifier connected to V_1 and V_2.

Further developments enabled silicon crystals of even higher resistance to be made. This was achieved by diffusing lithium into the material, creating a detection-sensitive layer 5 to 10 mm thick. This was sufficient for measuring X-radiation and high-energy particles, e.g. deuterons at about 25 MeV [5]. If germanium is used instead of silicon the same technology can be used for making detectors for measuring gamma radiation. These germanium detectors gave a previously unattainable energy resolution for gamma radiation (*fig. 2*).

As well as being used to detect radiation and particles for nuclear research, semiconductor detectors are also employed in many other kinds of investigation, such as materials analysis — with neutron beams, ion beams and X-radiation — and medical research — with gamma-ray cameras and for dosimetry.

The studies made by the Philips group have mainly been concerned with the manufacture of detectors with highly reproducible characteristics. We shall now discuss some of the results in more detail.

Dr Ir W. K. Hofker is with Philips Research Laboratories, Amsterdam Department.

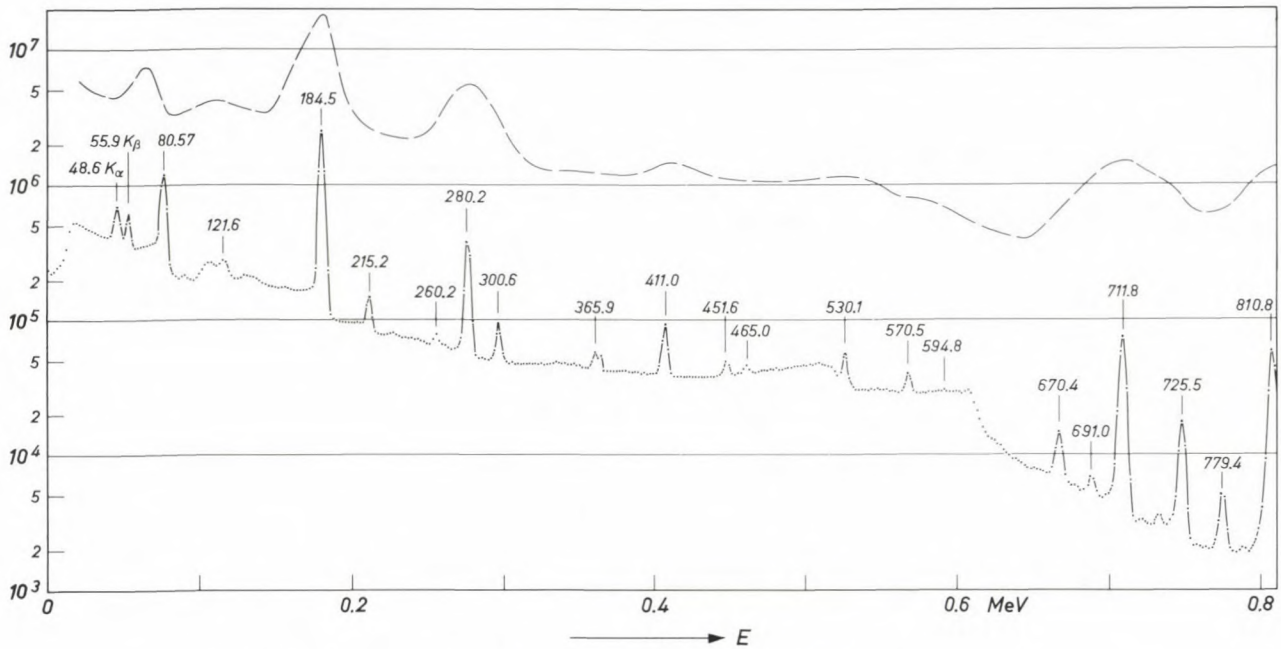


Fig. 2. Part of a gamma spectrum — of holmium 166 m — recorded by a germanium detector. Each dot represents the number of pulses N in an energy range of 2 keV as a function of the energy E. A very large number of separate lines can be seen: the quantum energy is given for most of them. The dashed line at the top shows the same spectrum, measured with an NaI(Tl) scintillation detector. (For clarity this spectrum has been shifted upwards in the figure.) The data were made available by Dr Ir J. Konijn of IKO.

Results from our investigations

Surface-barrier detector

For surface-barrier detectors [6] we have developed a method for treating a surface of a few square centimetres. The detectors produced were unaffected by a reverse voltage of more than 100 V. The reverse current — which is responsible for the noise — was kept to below 1 μA. Unlike most of the detectors that were available at the time, surface-barrier detectors were suitable for use in a high vacuum. (In most experiments the particles have to be observed in a vacuum.) Our technology also enabled a high output to be obtained in quantity production.

Energy-loss detector

In a surface-barrier detector only part of the thickness between the front and the back of the detector is sensitive enough to detect a particle. In the remainder of the thickness the particle does in fact lose energy,

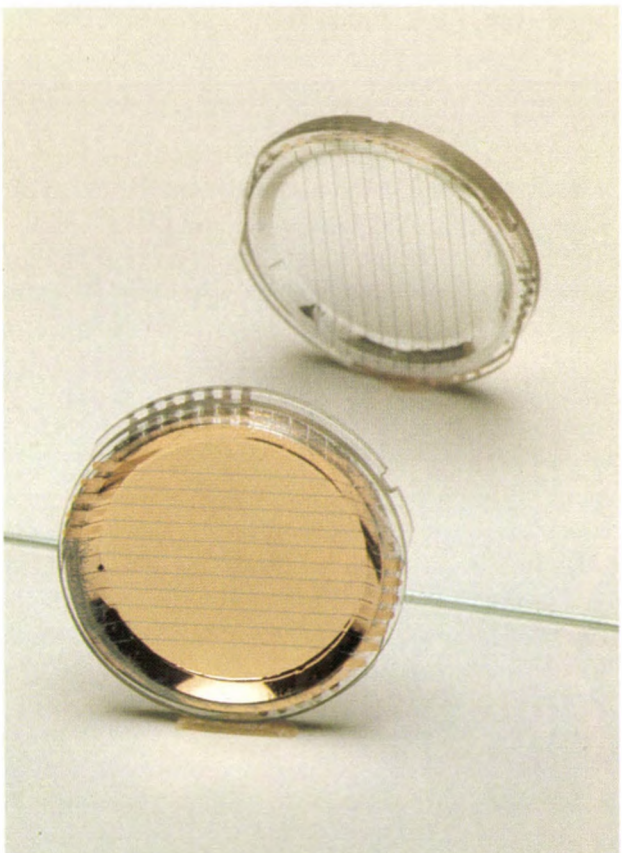


Fig. 3. Checkerboard detector, with a set of electrodes in the form of strips on the front and back. The strips at the front are at right angles to those at the back (see mirror image).

[1] See for example W. K. Hofker, Geiger-Müller counters, this issue, page 296.
[2] P. J. van Heerden, The crystal counter, Thesis, Utrecht 1945. P. J. van Heerden, Physica **16**, 505, 1950. P. J. van Heerden and J. M. W. Milatz, Physica **16**, 517, 1950.
[3] K. G. McKay, Phys. Rev. **76**, 1537, 1949.
[4] J. Mayer and B. Gossick, Rev. sci. Instr. **27**, 407, 1956.
[5] More information on Si(Li) and Ge(Li) detectors can be found for example in W. K. Hofker, Semiconductor detectors for ionizing radiation, Philips tech. Rev. **27**, 323-336, 1966.
[6] See for example W. K. Hofker, K. Nienhuis and J. C. Post, α-particle spectrometry with semiconductor detectors, Philips tech. Rev. **30**, 13-22, 1969.

but it is impossible to convert this energy loss into an electrical signal. In certain experiments, however, it is necessary to place a 'thick' detector behind this 'thin' detector. The identity of the particle can then be determined. A particle loses part of its energy in the thin detector, and the remainder in the thick detector, and the product of the two differs for different types of particle. However, it is then necessary to be able to measure the entire energy loss in the thin detector, which is then called an energy-loss detector.

With the same technology that we had used for making surface-barrier detectors we have also made energy-loss detectors. In these detectors an aluminium contact layer is applied to the thin oxide layer at the rear. In spite of the high field-strength at this contact the reverse current, and hence the noise, have been found to be very small.

Checkerboard detector

The principle of the energy-loss detector was used to develop the 'checkerboard detector' [7]. Instead of continuous contact surfaces this detector has contact strips on both sides. The strips on one side are perpendicular to those on the other side, hence the name (*fig. 3*). A captured particle induces an electrical signal in one of the strips on either side of the detector, so that the position where the particle enters the detector is within the 'field' where the activated strips intersect. A checkerboard detector therefore effectively consists of about 90 fields. In the BOL [8] detector telescopes (*fig. 4*) the checkerboard detector — thickness about 0.3 mm — was used as an energy-loss detector. Compared with a conventional energy-loss detector a checkerboard detector has ten times the angular resolution for the same value of the total area, i.e. for the same solid angle.

Reliability

In a BOL detector telescope [8] the checkerboard detector was combined with one or more Si(Li) detectors [5] — each about 5 mm thick — which were also developed by the Philips group at IKO. A total of about 100 semiconductor detectors were used in the BOL instrument. For the proper functioning of BOL it was therefore extremely important that the detectors should be very reliable.

Even stricter requirements for reliability are the rule in space studies. Several space laboratories have selected Elcoma surface-barrier detectors for their projects. These detectors were manufactured with the technology developed by the Philips group at the Institute. One of the users is the Utrecht Laboratory for Space Research, who are using the detectors to measure the energies of protons in the NASA satellite ISEE-3, launched on 12th August 1978. These satellite studies are being made to try to track down sources of proton radiation in interplanetary space to find out if there is any correlation with solar flares. The satellite experiment includes three telescopes, each equipped with a Philips detector (*fig. 5*) [9]. The experiment is expected to continue for several years.

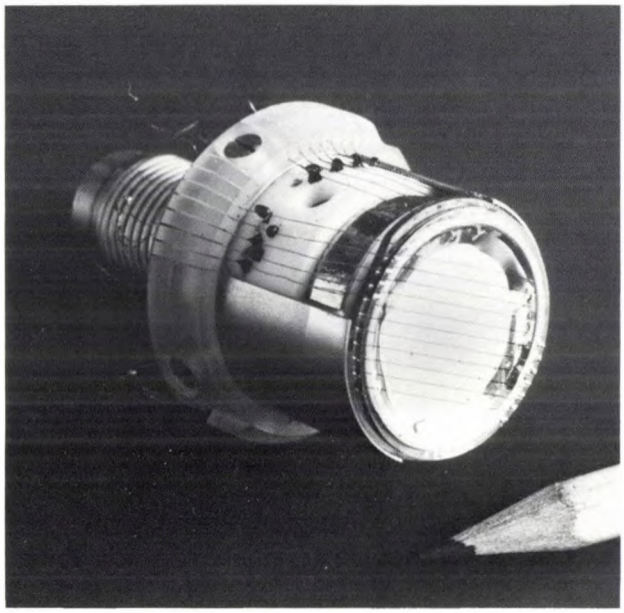


Fig. 4. Detector telescope for BOL. The shiny patch behind the checkerboard detector is the surface of an Si(Li) detector. The wires provide the electrical connections from the checkerboard strips to amplifiers (not shown).

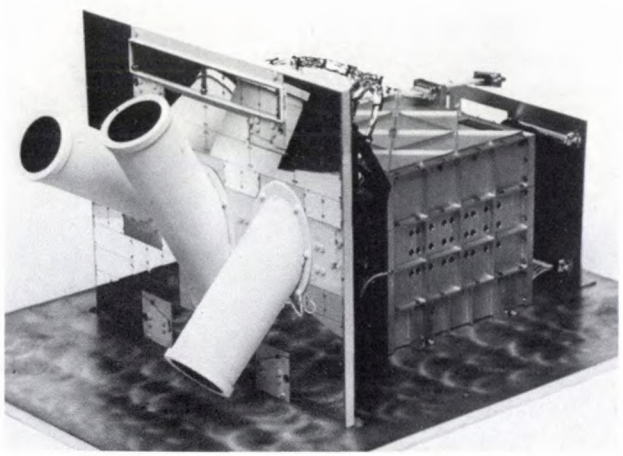
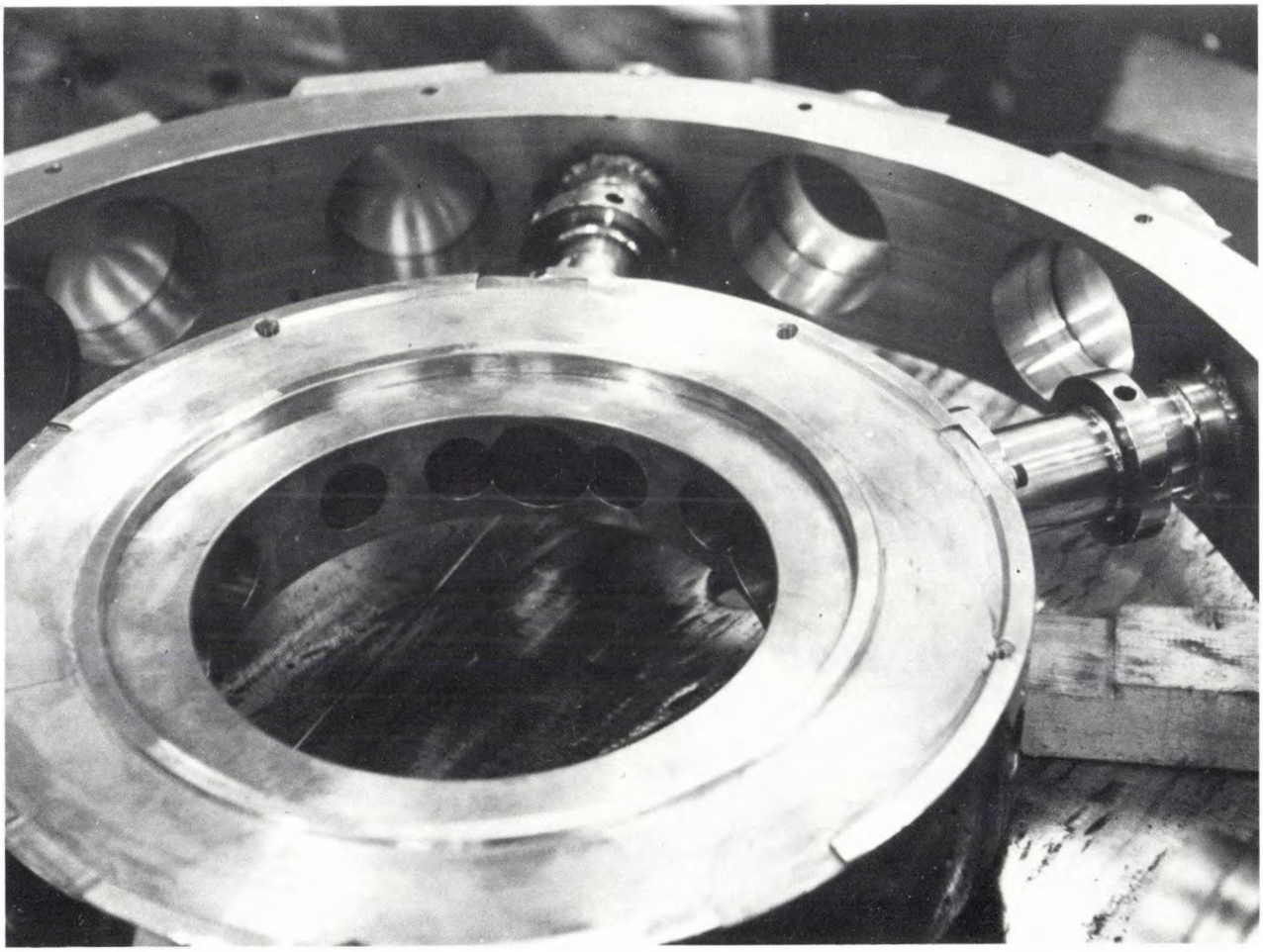


Fig. 5. Flight model of the proton spectrometer, now mounted in the NASA satellite ISEE-3. The instrument has three telescopes, each with detection equipment for measuring the energy and direction distribution of proton fluxes in interplanetary space. Each of the telescopes has a Philips surface-barrier detector for determining the proton energy.

[7] W. K. Hofker *et al.*, IEEE Trans. **NS-13**, No. 3, 208, 1966.
[8] See for example R. van Dantzig, BOL, this issue, page 302.
[9] J. J. van Rooijen, G. J. van Dijen, H. Th. Lafleur and P. Lowes, Space Sci. Instr. **4**, 373, 1979.



The inner ring of the BOL scattering chamber. The hemispheres were attached to the ring, which also carried a mechanism for bringing different targets into the external beam of the synchrocyclotron.

BOL

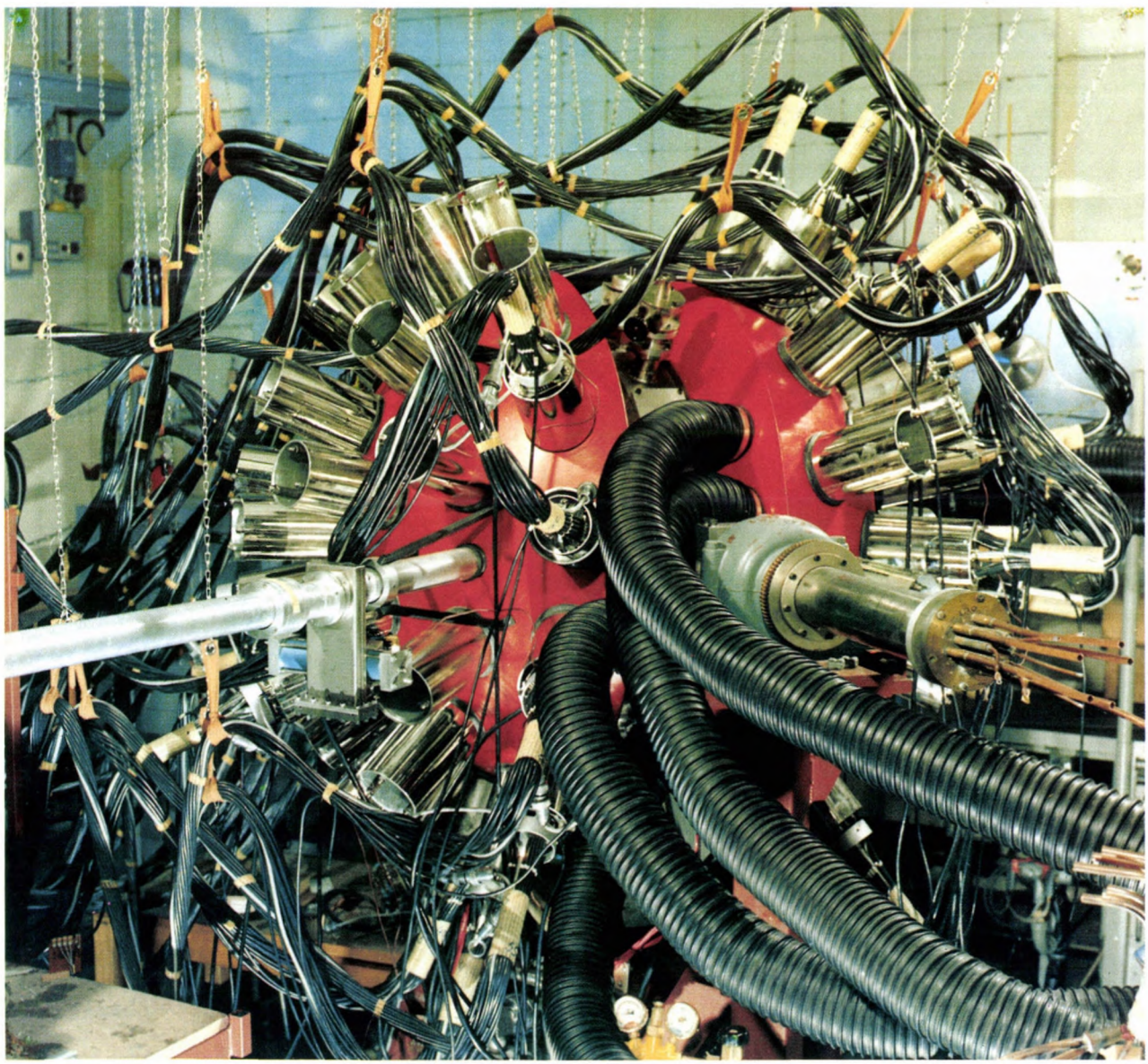
R. van Dantzig

The 'BOL' measurement system for nuclear physics (*fig. 1*) [1][*] was built by IKO during the period 1965 to 1969; it was called BOL because of its spherical shape (*'Bol'* is the Dutch word for sphere). The equipment and instruments were designed for studying nuclear reactions induced by the external particle beam of the synchrocyclotron [2]. The system was particularly suitable for studying nuclear collisions in which two or more of the escaping nuclear particles had to be detected so as to reconstruct the collision process.

BOL was the brainchild of Dr L. A. C. Koerts — who was later to join the Philips staff; the system

Dr R. van Dantzig is with IKO (Institute for Nuclear Physics Research), Amsterdam.

Fig. 1. The 'BOL' measurement system for nuclear reactions. Particles accelerated in the synchrocyclotron are guided through a pipe — on the left — into the spherical scattering chamber, where the particles cause nuclear reactions in a target. Reaction products are detected with 64 detector telescopes; the connecting wires and cables for these can be seen in the figure. The black hoses are part of the air-cooling system for the electronic circuits outside the vacuum.



was also built under his supervision. The Institute and Philips worked closely together on this project, particularly with regard to the particle detectors [3], which formed an essential part of the system. 64 detector telescopes (*fig. 2*) — each containing one of the 0.3 mm thick checkerboard detectors developed for BOL and one or more Si(Li) detectors of 4 to 5 mm thickness — were mounted in a spherical scattering chamber, at about 10 cm from the centre, where a target was located. 13% of the surface of the sphere was 'covered' by detector devices, thus ensuring a particularly high detection yield. The detection system has been compared by Koerts to a giant, inward-turned insect's eye.

incidence on the checkerboard detector, coincidence relationships, and dead time) for each detected particle were stored on magnetic tape with the aid of a DEC PDP8 process computer. For complicated 'on-line' calculations the data could also be passed directly to an Electrologica EL-X8 computer and presented on a PDP8 display. The interface equipment between the computers was developed in close cooperation with Philips and Electrologica.

This collection of almost autonomous pieces of equipment could operate automatically as a single unit because of the development of suitable software. For example, for asynchronous processing of data originating from the detection equipment and for

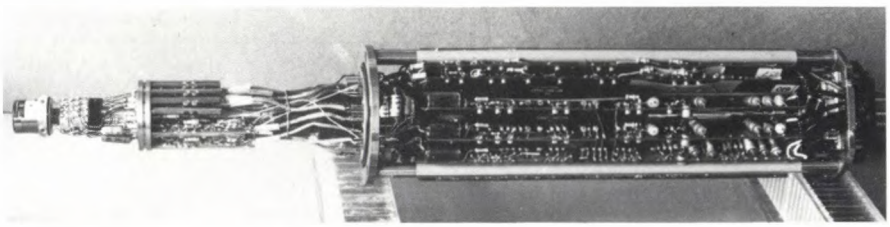


Fig. 2. A BOL detector telescope. The checkerboard detector and Si(Li) detector in the head of the telescope can be seen on the extreme left, with the charge amplifiers connected to them. The detectors and charge amplifiers were located inside the vacuum of the scattering chamber.

The detector telescopes were held in place by a spherical copper shell about 20 cm thick, in which the positioning holes had been very accurately located (*fig. 3*) [4]. The purpose of this 'inner sphere' was to cool the detectors (-20 °C) and hence reduce the electrical noise. Amplifiers were mounted as close to the detectors as possible. They amplified the weak detector signals to make them suitable for transmission over more than 1000 coaxial cables to the electronic measurement and control circuits, which contained about 100 000 discrete components. (The electronic circuits were developed at the Institute under the supervision of R. F. Rumphorst who was also later to join the Philips staff.) The inner sphere, the detector telescopes and the preamplifiers were located in a vacuum (10^{-4} Pa), enclosed by the outer sphere. The measurement and control circuits received data from the telescopes relating to: the energy, the point of incidence (and hence the direction of movement) and the nature of every particle detected, and coincidences with events detected in other telescopes.

The energy of a detected particle could be rapidly digitized, to an accuracy of 0.01%, in each of the 64 electronic detector channels. 72 bits of information (relating to aspects such as energy loss and point of

interfacing the computers the PDP8 was provided with a multi-task monitor and 'drivers' for the various data streams. A high-level programming language (SIMPLEX) was designed and implemented for the EL-X8. This language met the requirements of the data processing and was based on a standardization of the data streams ('window-processing'). Processed data, which often had to be studied in multi-dimensional spaces, could be presented on a purpose-built graphic display; here again Philips Research Laboratories and Electrologica had an important part to play.

Much of the work associated with the building of BOL lay on the fringe of what was technically feasible. This was a great stimulus for the many people who applied themselves so enthusiastically to making the instrument a practical reality.

[*] The author would be pleased to supply a complete list of references relating to the construction and the research.

[1] L. A. Ch. Koerts, K. Mulder, J. E. J. Oberski and R. van Dantzig, The "BOL" nuclear research project, Nucl. Instr. Meth. **92**, 157-160, 1971.

[2] See for example G. Luijckx, The cyclotron, this issue, page 290.

[3] See for example W. K. Hofker, Semiconductor detectors, this issue, page 298.

[4] K. Mulder, R. van Dantzig, J. E. J. Oberski and L. A. Ch. Koerts, Mikroniek **9**, 240, 1969 (in Dutch).

Nuclear-physics research

For studying nuclear processes particles accelerated in the synchrocyclotron were allowed to collide with stationary atomic nuclei in a target. In such a collision various processes can occur. The simplest of these is elastic scattering, rather like a collision between two billiard balls. Inelastic collisions can also occur. In these collisions kinetic energy may be converted into mass or into internal excitation energy, or vice versa; one of the colliding particles may also become unstable. After some time such a particle may return to a stable state by emitting electromagnetic radiation or, if the energy of impact is high enough, it may emit one or more of the nuclear particles from which it is built. It is also possible that during the actual collision one of the colliding particles may break up immediately into fragments. The dynamics of all these processes is extraordinarily complicated, particularly if the colliding atomic nuclei consist of many nucleons (i.e. protons and neutrons). In the research carried out using BOL, attention was directed mainly towards reactions in which only a few nucleons have a part to play or in which a group of nucleons could be considered in practice as a single entity, a 'cluster'. For such reactions it was possible to determine how energy and momentum were distributed among the various escaping particles. Because of the large number of detectors and the ability of the checkerboard detectors to determine direction it was possible to cover the great majority of all the possible final momentum states — determined by four independent degrees of freedom in the four-dimensional kinematic phase space. For a given nuclear reaction this makes it possible to determine the probability, expressed as the 'differential cross-section', for different kinematic final states.

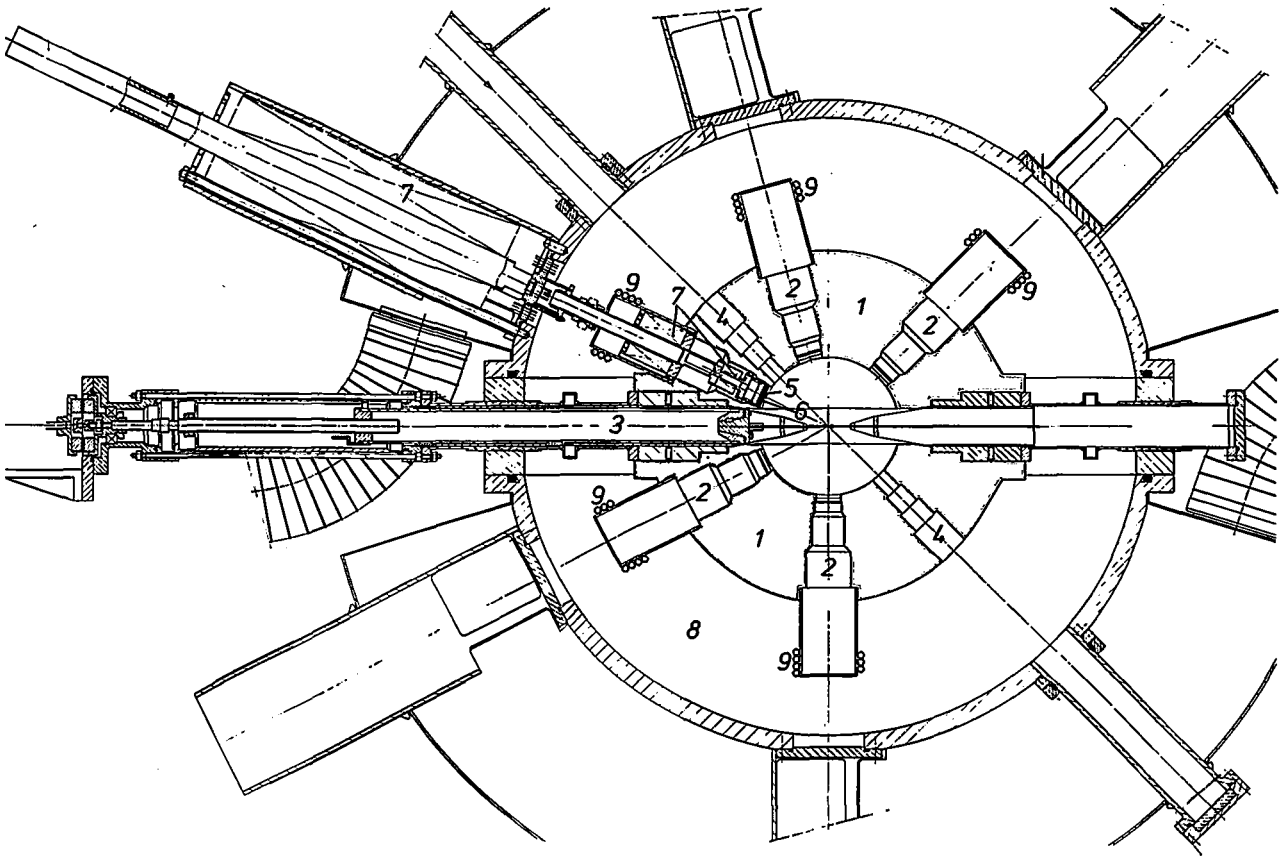


Fig. 3. Diagram of the cross-section of the BOL scattering chamber. Holes have been accurately machined in the spherical copper shell *1* for the detector telescopes *2*. The target transport system *3* could be used to introduce different targets into the beam, which was guided through the holes *4*. Checkerboard detector *5*, Si(Li) detector *6* and charge amplifiers *7* were mounted in the vacuum *8*. The spherical shell and detector telescopes were cooled by Freon *9*.

Theoretical interpretation

Some examples of experimental results obtained with BOL will be discussed later. Two important aids to analysis and interpretation are firstly, the 'exact' theory of three-particle processes introduced by the Russian mathematician L. D. Faddeev and secondly two simplified models that can be used to obtain an explanation of the most important collision phenomena that is surprisingly good in view of the simplifications.

Quasi-Free Scattering (QFS), for example, is one of these phenomena that can be described with the aid of a simplified model, the 'impulse approximation', as a collision between one of the two atomic nuclei as a single entity and one of the constituents of the other atomic nucleus. In this model the remaining constituents only betray their presence through the way they affect the momentum of the colliding particle.

The second model is that of Final-State Interaction (FSI), in which some particles that emerge from the primary region of the collision remain so close together for an appreciable time that they can be considered as a single unit in a description of the primary collision process. A mathematical description of this model has been given by K. M. Watson and A. B. Migdal.

Complete measurements

In the measurements made using BOL special emphasis was placed on the processes in which the momentum, energy and identity of all escaping particles could be reconstructed. Such measurements are said to be *kinematically* complete. The checkerboard structure was absolutely essential if we were to combine accurate determination of the momentum direction with a high detection yield. Almost all the final states that occur in a collision were recorded *simultaneously* with BOL; in this respect again the measurements can therefore be designated as virtually complete. Because of the large number of detector telescopes, information could be obtained on processes occurring in practically the *entire* kinematic *phase space*, so that the measurements were also complete in this aspect.

Analysis of the data

In a typical experiment several hundred magnetic tapes were filled with data in about a hundred hours. The data was then analysed from various points of view. The number of detected events as a function of the energies of detected particles — energy spectra — was determined for specific angles, while on the other hand the angular dependence was studied for specific values of the energy. The results of such an analysis were compared with theoretical calculations in which Monte Carlo simulation techniques had been used. This enabled an experiment to be simulated on the computer in accordance with a theoretical model, in which due allowance was taken of the particular circumstances of the experiment, including the experimental limitations. The measured results, in conjunction with the theoretical analysis, often led to conclusions that would have escaped if we had used a conventional experimental arrangement with only a few detectors.

The studies made with BOL have so far led to six dissertations, fifteen theses and dozens of publications and conference contributions. Other publications are being prepared.

Examples of research results

Reactions with three nucleons

The collision process between a proton (p) and a deuteron (d), in which two protons and a neutron (n) emerge (p + d → ppn), was studied for two collision energies. First, the regions of the kinematic phase space in which the quasi two-particle processes QFS and FSI predominate were analysed. This means that in the computer analysis of the data kinematic conditions were selected in which mainly QFS and FSI could be observed (*fig. 4*)[5]. Secondly, as the analysis developed further, the process was studied in the kinematic regions that are remote from the regions where QFS and FSI predominate. Very few measured results were available at the time for these regions, where in comparisons with numerical calculations the greatest sensitivity to oversimplification of the nucleon-nucleon potential might in fact be expected. Finally, we made an analysis for almost the entire phase space, with measured results reproduced in four-dimensional 'arrays' corresponding to the four kinematic degrees of freedom of the reaction.

There was some surprisingly good agreement between the measured results obtained with BOL and the theoretical computations based on Faddeev's scheme, carried out by Utrecht theoreticians[6], but there were also some very marked differences. It is not yet clear whether these differences are primarily a consequence of terms in the nucleon-nucleon potential that were neglected in the calculations, or whether the differences are more fundamental. We hope that further calculations will provide a more detailed answer to this question.

[5] B. J. Wielinga *et al.*, Lett. Nuovo Cim. **11**, 655, 1974, and Nucl. Phys. **A261**, 13, 1976.

[6] W. M. Kloet and J. A. Tjon, Nucl. Phys. **A210**, 380, 1973.

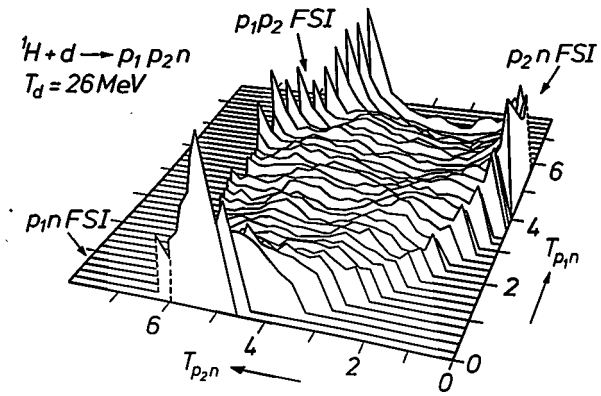


Fig. 4. Isometric representation of the results of a measurement made with BOL. Protons were bombarded with deuterons at an energy of 26 MeV. Two protons p_1, p_2 and a neutron n emerged from the nuclear reaction. Both protons were observed in coincidence over almost the entire range of angle and energy; the angle and energy of the neutron could then be calculated. The observations are represented here as a 'Dalitz plot': the number of events as a function of the relative energy between one proton and the neutron (T_{p1n}) and between the other proton and the neutron (T_{p2n}), integrated over the two remaining degrees of freedom in phase space. The high peaks in the figure correspond to FSI processes: final-state interaction for a low relative energy between proton and neutron ($T_{p1n} \approx 0$ or $T_{p2n} \approx 0$) and between proton and proton ($T_{p1n} + T_{p2n} \approx$ maximum, i.e. $T_{p1p2} \approx 0$). QFS processes appear in the figure as raised bands along lines of constant relative energy ($T_{p1n} = T_{p2n} = T_{p1p2} \approx 4.5$ MeV).

Reactions with four nucleons

In the case of three nucleons it was possible to use the 'exact' method of calculation based on Faddeev's scheme to make a quantitative comparison between a theoretical prediction and an experimental result, but this cannot yet be done in the case where there are four nucleons. For the present, analyses of reactions with four nucleons are therefore mainly intended to give a qualitative grasp of the phenomena, preferably independently of any theoretical models. In addition, phenomenological models are used, such as the QFS, FSI and 'transfer' models. The transfer models describe the exchange of a constituent between colliding nuclei: 'pick-up' (from target nucleus to incident particle) and 'stripping' (vice versa).

In the reaction p + ^{3}He → dpp, for example, in which p-p and p-d pairs were detected with BOL in coincidence, it was found that some of the effects could be explained by assuming that the process takes place in two stages (a sequential model). On a more detailed examination, however, it was found that there were substantial deviations from this model as the result of contributions from other reaction mechanisms. We therefore investigated the relationship between different reaction mechanisms as well as their relative importance. In particular we tried to find out where these processes were located in the kinematic phase space and how far they extended.

In the reaction d + d → dpn we observed that quasi-free scattering between one deuteron and one nucleon of the other deuteron is an important feature over almost the entire phase space. The measured results were compared with theoretical calculations in which the most important reaction mechanisms were taken into account.

Another study of four nucleons was made with the aid of reactions on various atomic nuclei, which were induced by ^{3}He. This gave indications of a neutron pick-up process, in which an unstable excited state of ^{4}He nucleus is formed, which then decays into a proton and a triton or into two deuterons. The two end-products were detected in coincidence.

Disintegrating ^{3}He

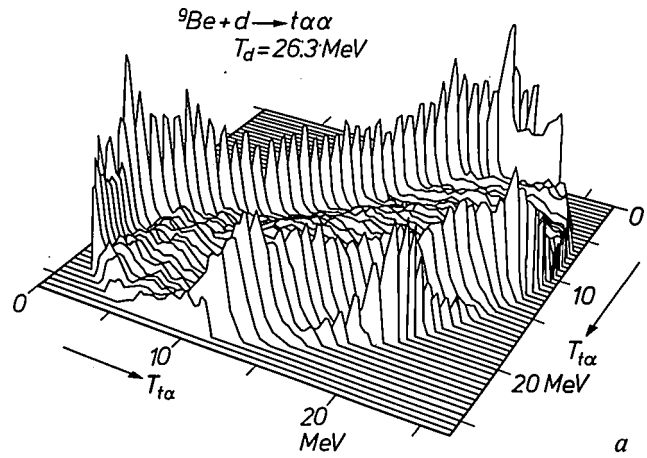
If a ^{3}He nucleus is allowed to collide with an atomic nucleus, the ^{3}He may disintegrate into a deuteron and a proton or into two protons and one neutron. With the BOL system the charged particles could be detected in coincidence. Correlations in energy and direction of the ^{3}He disintegration products were used to study the nature of the disintegration process and the state of the (heavy) residual nucleus after the collision, for gold, nickel, aluminium, carbon and beryllium nuclei. If the residual nucleus continued to remain in the ground state or in a low excited state the directional correlations showed that the ^{3}He had already disintegrated at the outer edge of the nucleus.

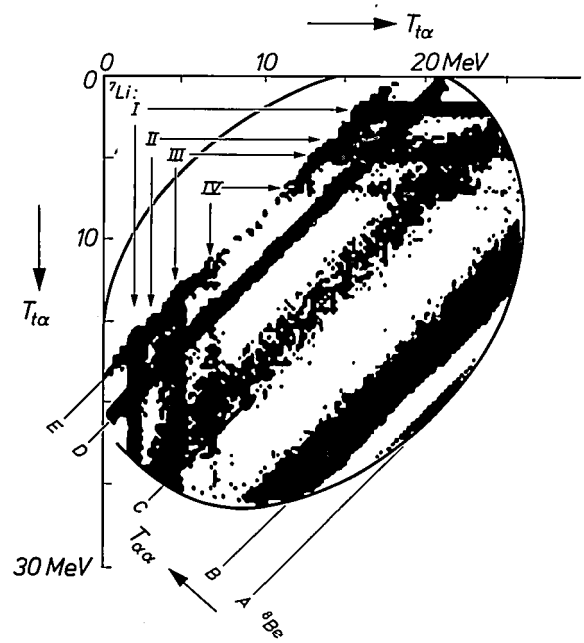
In this study proton pairs were also detected with strong FSI. These showed up in one detector telescope as 'double hits' and could be considered as a di-proton — i.e. an unstable ^{2}He nucleus — but because of the structure of the checkerboard detector they could still be analysed as a two-particle coincidence — but one with a small relative angle.

Reactions with three clusters

The description of reactions involving many nucleons can be simplified if most of the nucleons occur in clusters, which — during the process — can be considered to a good approximation as single particles. If moreover the reaction process takes place in two stages (the first stage is the formation of a metastable nuclear state with well-defined quantum numbers, the second is the decay of this state), detailed information can be obtained about the primary reaction process by observing the directional correlations of the decay products.

For example, inelastic scattering of deuterons was used to bring ^{12}C nuclei into an excited state at 9.3 MeV and the decay from this state into ^{4}He and ^{8}Be nuclei was then studied. The results were appreciably different from a theoretical analysis, using





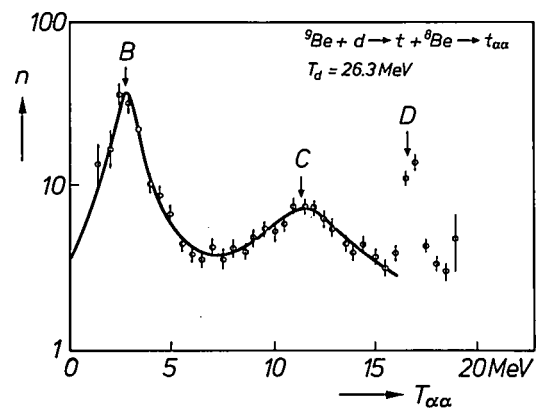


Fig. 5. Results of measurements made with BOL, in which ^{9}Be nuclei were bombarded with deuterons. The nuclear reaction may proceed via intermediate states (resonances) in ^{8}Be or via states in ^{7}Li, in both cases giving a triton (t) and two ^{4}He nuclei (α) as end-product clusters. *a*) Dalitz plot (for explanation see caption to figure 4). *b*) The Dalitz plot in (*a*) viewed from above. Here the bands A-E correspond to resonances of the intermediate nucleus ^{8}Be and bands I-IV correspond to resonances of the intermediate nucleus ^{7}Li. Where resonance bands intersect, the nuclear reaction proceeds simultaneously via different intermediate states. In the overlapping areas the interference of these different processes can be studied. In the areas of the Dalitz plot where there is no overlap a process via one of the two intermediate nuclei can be studied separately. *c*) Energy spectrum for the process taking place via the ^{8}Be intermediate nucleus (n number of events as a function of the relative energy $T_{\alpha\alpha}$). A similar spectrum can be given for the process that takes place via the ^{7}Li intermediate nucleus.

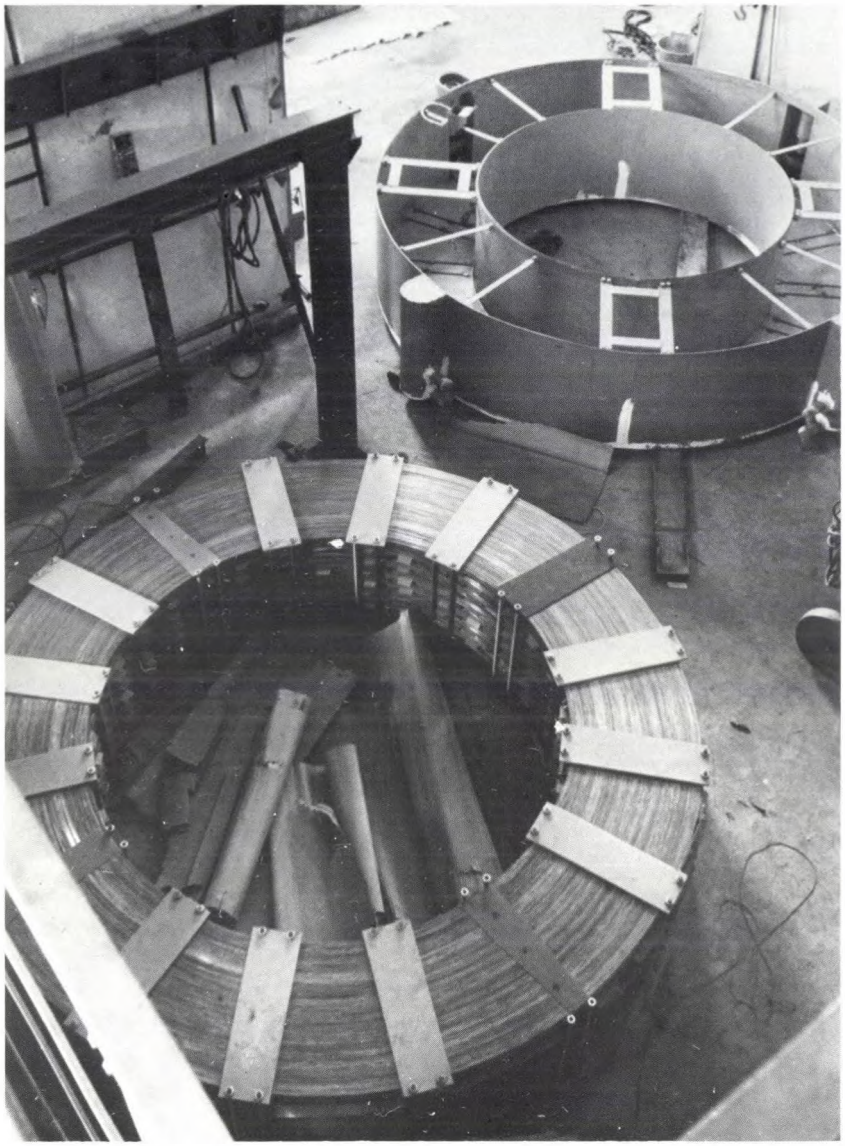
the 'Distorted Wave Model', made in a joint study with the Eindhoven University of Technology [7].

Studies were also made of reactions in which a deuteron collides with a ^{9}Be nucleus and can form, with the emission of a triton, an ^{8}Be nucleus in various excited states (*fig. 5*), which decay into two ^{4}He nuclei [8]. From the relations between the energies and directions of the triton and the two ^{4}He nuclei we obtained a great deal of information about the excited ^{8}Be intermediate states. The contributions from various reaction mechanisms were analysed, with particular attention to the relationship between them. For example, we made measurements on the (d,t) reaction, in which the deuteron picks up a neutron from the ^{9}Be nucleus, the residual nucleus ^{8}Be then decaying into two ^{4}He nuclei, and compared it with the (d,^{4}He) reaction, in which the incident deuteron picks up a deuteron from the ^{9}Be nucleus, forming an excited intermediate state in ^{7}Li that decays into a triton and a ^{4}He nucleus. The end-products are the same in both cases, so that the two processes are quantum-mechanically indistinguishable and therefore interfere. So as to compare the results of this investigation, corresponding reaction processes were also studied using reactions of protons with ^{9}Be.

Many other reaction processes have been studied with the help of BOL. One example is the reaction of deuterons with ^{9}Be, which produces a proton, a ^{4}He nucleus and a ^{5}He nucleus. This particular study revealed a previously unknown state in ^{10}Be.

[7] B. J. Verhaar, W. C. Hermans and J. E. J. Oberski, Nucl. Phys. **A195**, 379, 1972.

[8] L. R. Dodd, M. A. A. Sonnemans and R. van Dantzig, Lett. Nuovo Cim. **12**, 597, 1975; see also M. A. A. Sonnemans, J. C. Waal and R. van Dantzig, Phys. Rev. Lett. **31**, 1359, 1973.



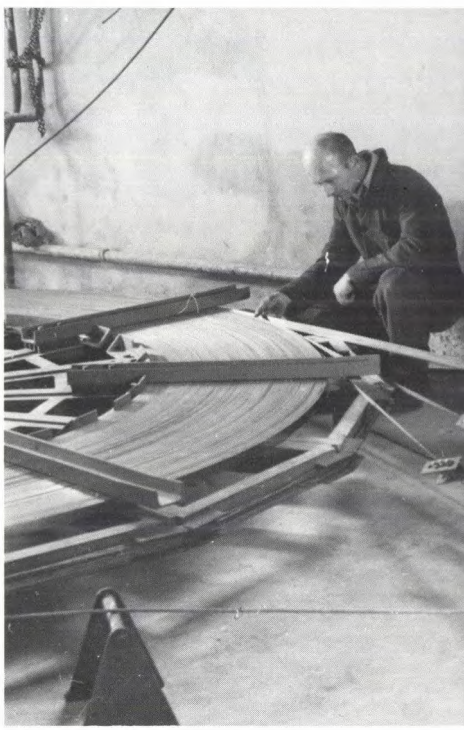
Some pictures taken during the construction of the Philips synchrocyclotron:

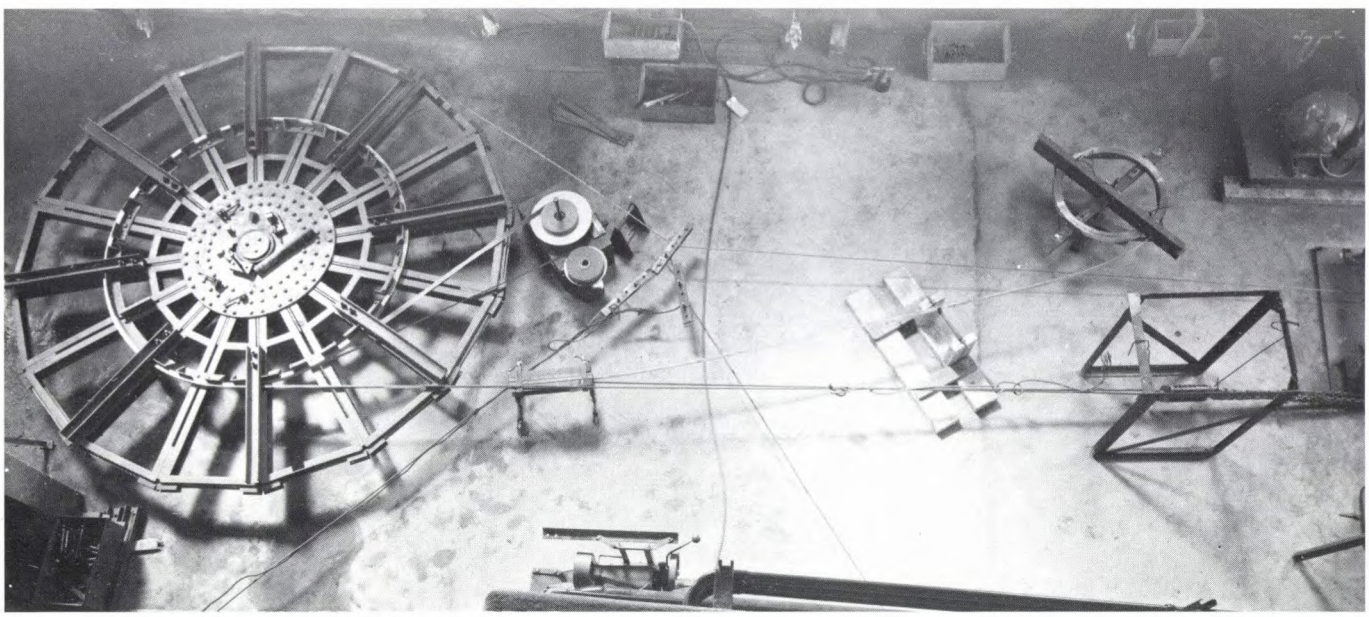
Winding the magnet coils,

Delivering some of the magnet iron,

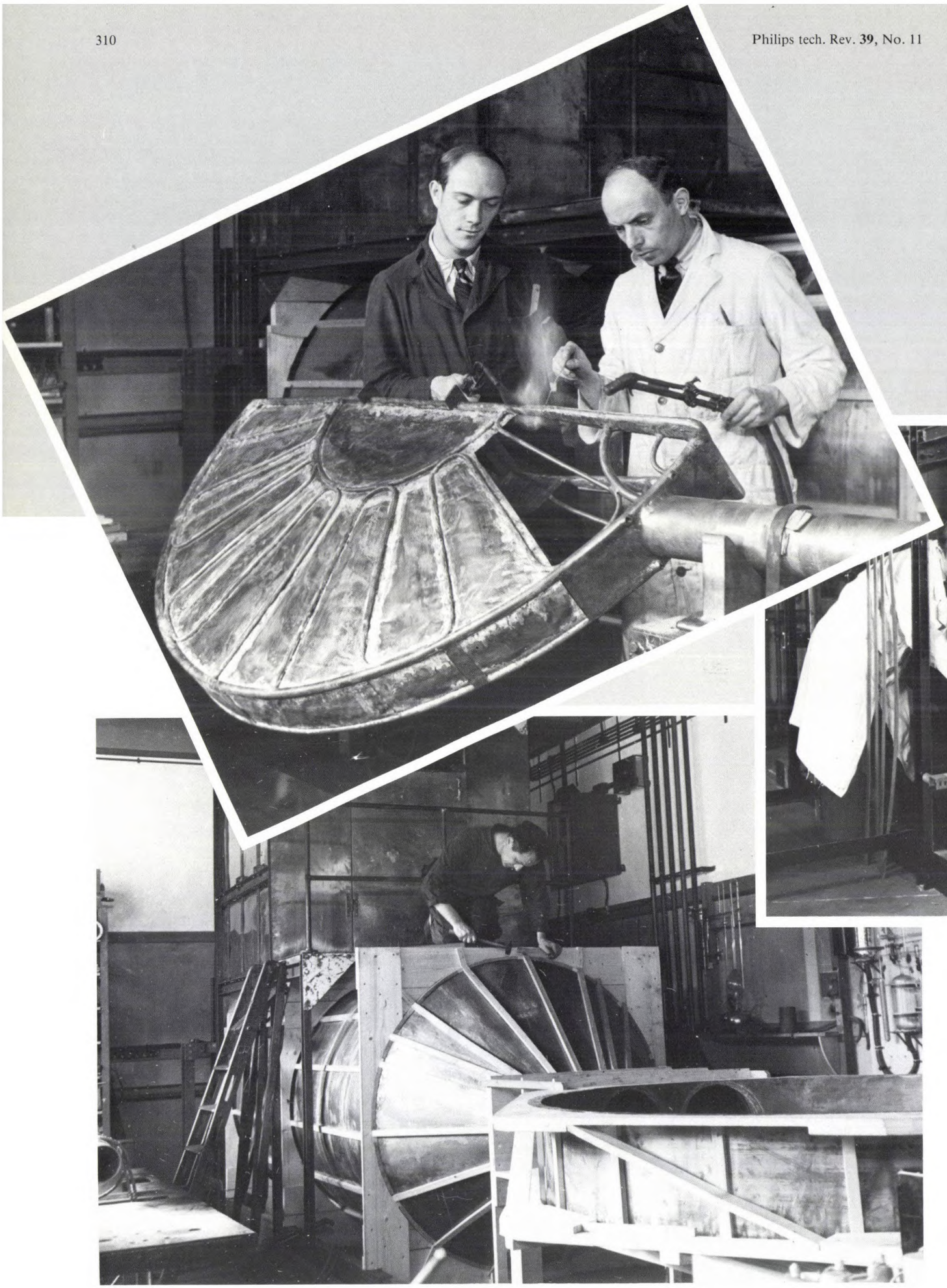
The dees,

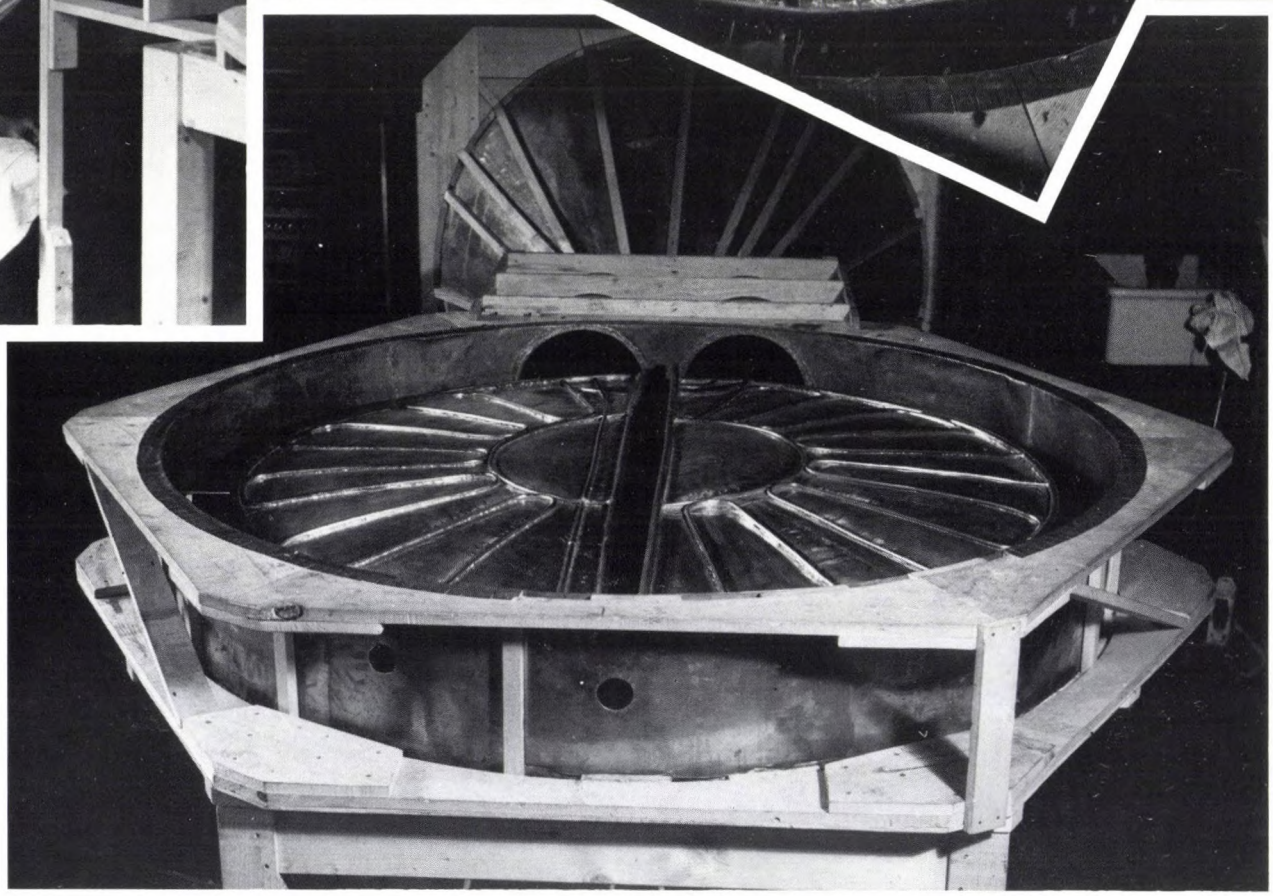
Part of the radio-frequency system.



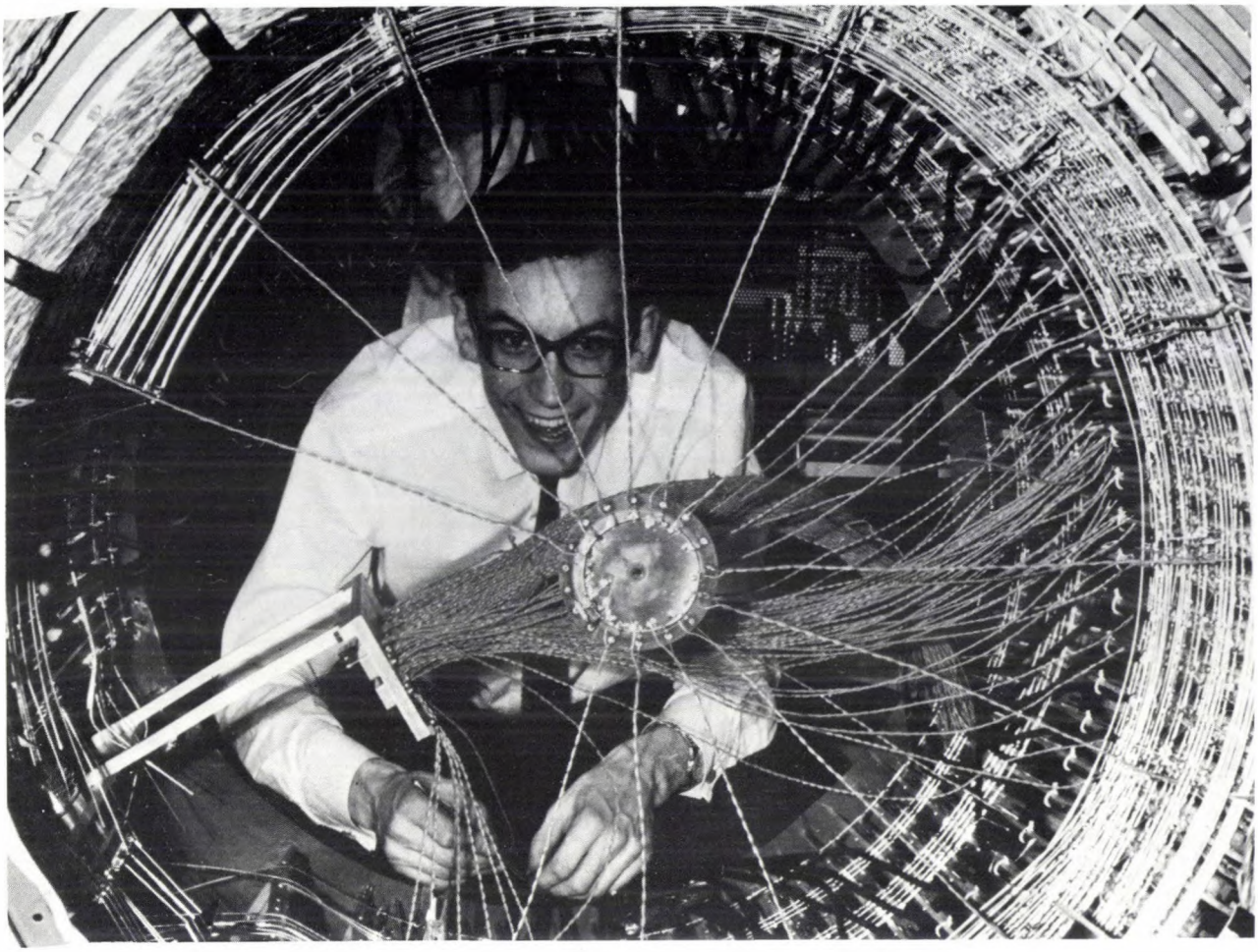


TIC. TRANSPORT





Philips tech. Rev. **39**, 312-314, 1980, No. 11



Assembling the wiring in the cylinder. There are 2 × 64 printed-circuit boards with electronic circuits for the BOL system.

Electronics for nuclear physics

W. K. Hofker

As a result of their work on radiation detectors, the Philips group at IKO found that they had to pay more and more attention to 'electronics for nuclear physics', i.e. the design of the electronic circuits for processing the signals originating from these radiation detectors.

In about 1950, when the signals to be processed came from Geiger-Müller counters, the electronic circuits were still quite simple, even for that time. The pulses only had to be counted [1]; their magnitudes did not have to be measured. Even then, however, it was important to ensure the smallest possible dead time, i.e. the time immediately following a signal during which the instrument cannot handle another signal. However, the dead time is determined both by the dead time of the counter tube and by the dead time of the electronic circuit, and optimization of these dead times was another important subject of study at the time [2].

Dr Ir W. K. Hofker is with Philips Research Laboratories, Amsterdam Department.

Spectrometry with semiconductor detectors

After the Geiger-Müller counters and the scintillation counters, semiconductor detectors began to appear in about 1956 [3]. A semiconductor detector, unlike a Geiger-Müller counter, can be used to determine the energy of a captured particle (e.g. a helium nucleus or an electron) or a photon (e.g. X-ray and gamma-ray quanta). This is because the quantity of charge that the detector gives up to the electronic circuit is proportional to the energy transferred from the captured particle to the detector material. Once an energy spectrum is known, e.g. for a radioactive specimen, it is possible to find out the nature of the energy-level diagram for that radioactive atomic nucleus. With semiconductor detectors the energy resolution could be greatly increased in comparison with results obtained earlier with proportional counters or scintillation counters. Semiconductor detectors are also widely used in analysis of the composition and structure of materials.

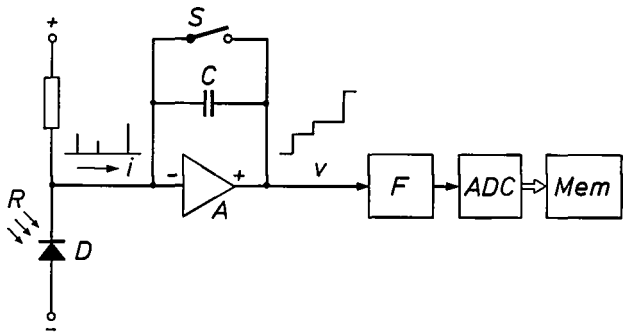


Fig. 1. Block diagram of an energy spectrometer. Radiation quanta *R* liberate the charge in the detector *D*; this charge is collected and integrated in the charge amplifier *A*, which provides a stepped signal. *C* integrating capacitance. *S* discharge switch. The filter *F* improves the signal-to-noise ratio. The analog-to-digital converter *ADC* then converts the amplitude of the signal into a digital value. The quanta are classified in the memory *Mem* by their energy and are counted.

Fig. 1 is a schematic representation of the basic structure of a nuclear spectrometer with a semiconductor detector. The detector is connected to a charge amplifier, i.e. an amplifier with capacitive feedback, which integrates the charge and gives a stepped output voltage. The magnitude of the step is proportional to the energy of an observed particle. The capacitance is periodically discharged by means of a field-effect transistor circuit, to prevent saturation of the amplifier. The output from the charge amplifier is passed through a bandpass filter and applied to an analog-to-digital converter (ADC) that converts the magnitude of the step into a binary number. The numbers are stored in a memory for further numerical processing.

Resolution

To take full advantage of the high resolution of the semiconductor detector the measurement accuracy of the entire spectrometer has to meet certain requirements: it must be better than 0.1%. The measurement accuracy can be affected by two kinds of perturbation. First of all, the measurement is affected by the electronic noise in the detector and the first stage of the amplifier. The bandpass filter is included to reduce the effect of the noise. The filter values are selected to give the maximum possible signal-to-noise ratio. This is obtained in theory if the filter gives an output signal that decays exponentially with time (see *fig. 2*). With such a curve, however, the signal has a relatively long trailing edge, and this can cause 'pile-up' of the signal: this 'pile-up' is the second major cause of perturbation. The differences in the arrival times of successive particles at the detector, for a given source intensity, are quite random, so that the time differences have a statistical distribution (a Poisson distribution) about a mean value. Consequently there is no periodicity in the time difference between successive signals from the detector, which means that the next signal may arise before the previous signal has finished, resulting in the signal pile-up just mentioned. The measurement of the amplitude of the signal and hence the determination of the energy of the captured particle are therefore perturbed and there is a reduction in the energy resolution. This effect increases as the mean count rate increases, so that the resolution required sets a limit to the maximum permissible count rate.

The Philips group at IKO have carried out detailed investigations to find out how to increase the signal-to-noise ratio and reduce the 'pile-up'. To illustrate

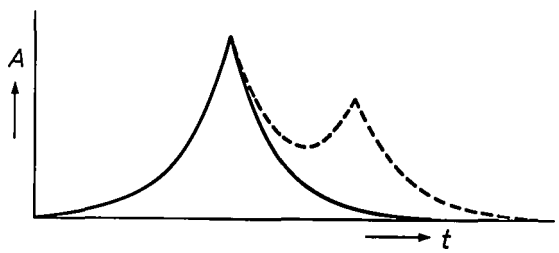


Fig. 2. Optimum filter response. The theoretically optimum filter gives an output signal with edges that vary exponentially with time ('cusp' curve, *solid line*). 'Pile-up' can easily occur because of the long response time. *A* signal amplitude, *t* time.

[1] See for example W. K. Hofker, Geiger-Müller counters, this issue, page 296.
[2] K. van Duuren, A. J. M. Jaspers and J. Hermsen, Nucleonics 17, No. 6, 86, 1959.
[3] See for example W. K. Hofker, Semiconductor detectors, this issue, page 298.

the results that the group has achieved, a brief description now follows of an instrument designed to give a good resolution of the semiconductor spectrometer even at high count rates.

Sampling ADC

In a circuit of the type developed by H. Koeman [4] the signal processing is different from that in a conventional spectrometer: the step signal from the charge amplifier is first of all rapidly digitized periodically (sampling period and conversion time are both 200 ns) and then filtered digitally. This is done by recording the values of the samples in binary numbers of 12 bits and storing these numbers in a shift register (*fig. 3*). The magnitude of the step, i.e. the actual measurement, is determined by comparing samples immediately before and immediately after the occurrence of the step. This comparison is initiated by means of a step-recognition circuit. The comparison is made for four samples before and four samples after the step. Various weighting factors can be assigned; for example the weighting factor may be made to increase as the sample becomes closer to the step. In theory, the best signal-to-noise ratio is obtained with weighting factors of 1 (farthest from the step), 1, 2, 4 (closest to the step). It has also been found, however, that an acceptable resolution can be obtained in practice with a simple average of the samples — i.e. identical weighting factors.

A 'pile-up' occurs with this circuit if two steps occur in the time taken for sampling a single step. In the circuit that we built this time is 1.6 μs, which gives a probability of pile-up of about a third of that in a conventional spectrometer.

As an example *fig. 4* illustrates a spectrum of radioactive iron measured with an Si(Li)-detector and a sampling ADC. The mean count rate was 300 000 events per second. There is still evidence of some 'pile-up', but the K_α and K_β X-ray peaks are clearly separated, whereas a conventional spectrometer would be saturated at such a count rate.

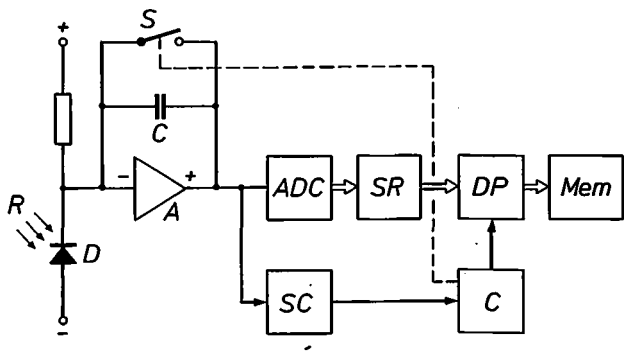


Fig. 3. Block diagram of an energy spectrometer on the Koeman principle. In this arrangement, unlike that of fig. 1, a sample of the output signal of a charge amplifier is periodically digitized (200 ns) by the analog-to-digital converter. The digitized samples are stored temporarily in a shift register, so that when the step-recognition circuit recognizes a signal the four preceding samples are still available. The digital processor calculates the step height from a total of eight samples, four preceding the recognition and four following it (digital filtering). The result is classified in the usual way in the spectrum memory and counted. *SR* shift register. *DP* digital processor. *SC* step-recognition circuit. *C* control circuit. Other symbols as in fig. 1.

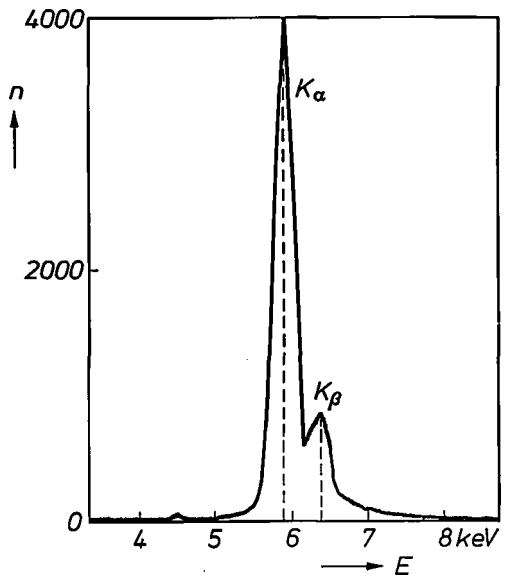


Fig. 4. Energy spectrum of ^{55}Fe for 300 000 events per second. The K_α and the K_β peaks are clearly separated. *n* number of counted pulses measured with a 4000 channel pulse-height analyser (one channel is equivalent to approximately 3.5 eV).

[4] H. Koeman, Nucl. Instr. Meth. 123, 161, 169 and 181, 1975.

Engineering technology

H. J. Akkerman

When work started on the construction of the cyclotron in 1946 there was a need for engineering support on the site. The postwar shortages meant that mechanical engineers had to manage as best they could with very few machines; the cyclotron project had one milling machine and a discarded pre-war lathe, and very few tools — a single calliper gauge, for example. A few heavy timbers were put together to make a work bench. In 1948, thanks to the American Marshall aid plan for Europe, a new lathe was acquired but it was a large British machine not well suited to our work. The most important components of the cyclotron, however, were made in the workshops of Philips Research Laboratories in Eindhoven, and the close cooperation that resulted from this situation was to have an important influence later on the development of the workshop at IKO.

In the fifties nuclear physics was a flourishing and expanding field of research, and IKO found themselves in increasing need of special engineering techniques so that they could construct better instruments for the work in hand. At that time, mechanical-engineering technology was also in a phase of rapid development and many new opportunities were being offered by the many new materials and processing methods becoming available. As a result the workshop grew into an engineering department with specialized sections for design and drawing, metrology, bonding techniques, vacuum and cryogenic engineering, electroplating and mechanical processes (*fig. 1*). The department could tackle other problem as well,

Fig. 1. General view of the workshop. In 35 years it grew from a little local workshop with four craftsmen to become a centre for engineering technology with a staff of 34.

H. J. Akkerman is with Philips Research Laboratories, Amsterdam Department, and is head of the Mechanical Technology Section at IKO.

This article arose partly out of discussions with J. J. Arendse (Philips Research Laboratories, Amsterdam Department) and G. C. Gerritsen (IKO).

because of the support it had in Eindhoven. This placed it in a position to carry out extensive projects such as the electron accelerators and 'BOL' for IKO, and the ion-implantation equipment for the Philips group at IKO.

A few examples will be quoted to give an impression of the kind of work that was carried out in the engineering department during the last 35 years.

Vacuum components

The vacuum system formed an important part of the cyclotron system [1]. The pressure in the vacuum chamber of the cyclotron (volume 800 l) had to be less than 10^{-4} Pa. When the cyclotron was in use there was a strong magnetic field, and the components also had to be able to stand up to the radioactive radiation released. In about 1946, when the cyclotron was being built, vacuum components for such applications were not an everyday item, and were not therefore commercially available.

This resulted in all kinds of standard designs being developed for vacuum components, including a design of vacuum valve that was unique at the time [2]. This valve had a toggle-lever construction enabling the pressure required to close a valve to be created by a relatively small force (*fig. 2*). A large number of these valves with orifice diameters between 50 and 350 mm have since been manufactured (*fig. 3*). Another then novel feature of the valves was that they could be operated either pneumatically or by hand. In the middle fifties pneumatic control systems had come into use in the Netherlands and a system of this type was used for remote control and automation of moving parts of the cyclotron.

Pneumatic automation was also used for the vacuum-lock system for the cyclotron target. In this system a target was introduced into the vacuum of the cyclotron and brought out again after bombardment, without altering the state of the vacuum in the chamber of the cyclotron in any way. The automatic system not only made it easier to change the target, it also made it safer. A short vacuum locking time was also important for specimens with a short half-life; the seven successive operations necessary were completed in eight seconds.

Cobalt needles

In 1969 about forty cobalt needles were manufactured for the former Philips-Roxane company. Each needle contained a small rod of radioactive cobalt. The needles were used in the treatment of cancer. The difficulty was that the needles had to be machined after

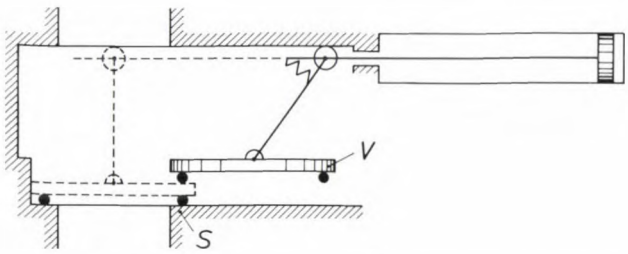


Fig. 2. Diagram of the vacuum valve developed in 1952. Only a relatively small force is necessary to press the valve *V* vertically down on to the seat *S*.

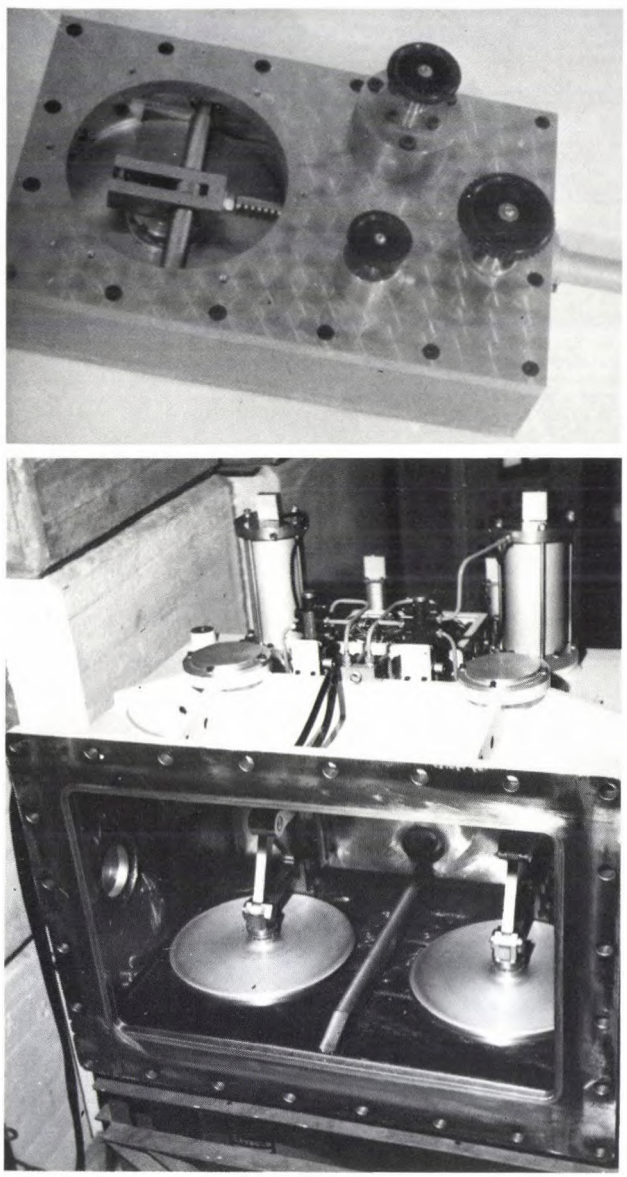


Fig. 3. Some examples of vacuum valves with orifice diameters between 8 and 35 cm. One of the features of these valves was that they could be operated either by hand or by compressed air.

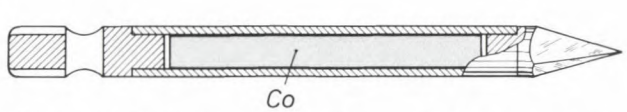


Fig. 4. Cobalt needle, diameter 1.1 mm, filled with radioactive cobalt (*Co*).

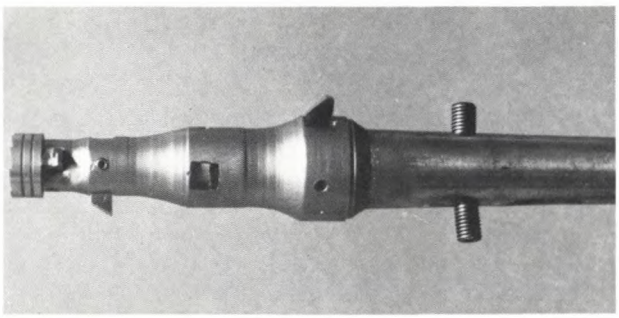


Fig. 5. Boring bar with six accurately positioned cutting tools. With this device the desired profile for a detector-telescope hole could be machined in a single operation. The profile of the hole can be seen in fig. 3 of the article on BOL, see page 304.

they had been filled with the radioactive material.

The needle was made from a stainless steel tube with a diameter of 1.1 mm (*fig. 4*). This was sealed off at one end by a clearly marked stopper with a nylon thread. The radioactive rod was then inserted and the other end was also sealed by a stopper, which was ground to form a three-cornered point. For safety reasons the joints had to be soldered with silver.

Various pieces of ancillary equipment were developed so that the filling, soldering and grinding operations could be carried out remotely at a safe distance.

BOL

In IKO's 'BOL' instrument for research in nuclear physics [3], 64 detector telescopes had to be accurately positioned. Each telescope had to be pointed at the centre of the sphere to an accuracy of 5 μm.

This was accomplished by drilling centring holes to the required accuracy in two hemispherical castings. Each casting was machined on a horizontal boring and milling machine. The workpiece was placed on a structure of three rotating worktables that could move perpendicular to one another. The point of rotation of this structure was set to the centre-line of the machine, using ball and feeler gauges to achieve the required accuracy. For the drilling a boring bar with six cutting tools (*fig. 5*) was used so that the entire profile of a telescope could be made in the workpiece in a single operation (*fig. 6*).

'Wire chambers'

Two 'wire chambers' have been constructed for the observation of particles in nuclear-physics experiments at IKO. A wire chamber operates in an analogous way to a Geiger-Müller counter, but contains a large number of anode wires, in this case as many as 1200 per chamber. The wires — of tungsten — have a diameter of 0.02 mm and are stretched across a frame

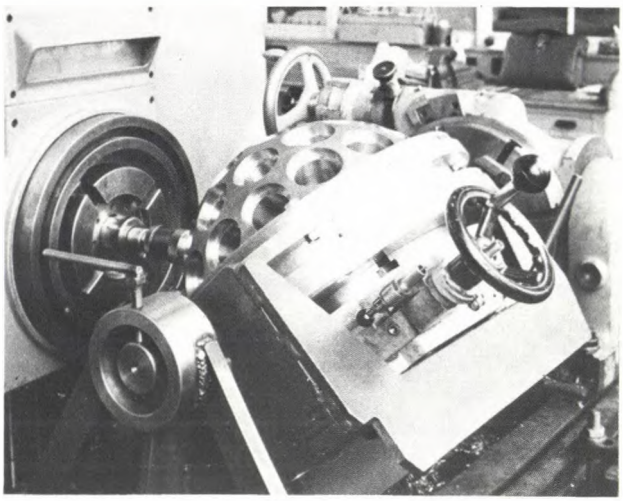


Fig. 6. Combination of three rotary tables mounted at right angles to one another for machining a hemisphere on a horizontal boring and milling machine.

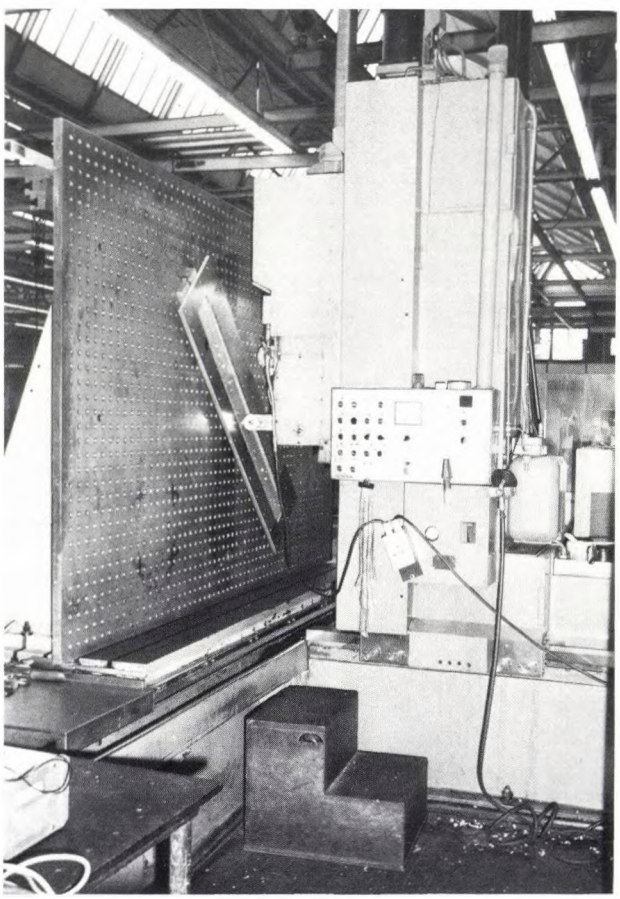


Fig. 7. Machining the locating grooves with a numerically controlled machine at the Philips Engineering Workshops in Eindhoven. A large number of grooves 70 μm deep are made in the strip, which is attached to a supporting plate about 2 metres above the ground (on the left in the figure).

[1] See for example G. Luijckx, The cyclotron, this issue, page 290.
[2] H. J. Akkerman, Fijntechniek **2**, 154, 1962 (in Dutch).
[3] See for example R. van Dantzig, BOL, this issue, page 302.

of dimensions 1220 × 110 mm. Two sheets of polyester film with a conductive coating are placed on either side of the wires and act as cathodes.

In the construction of the unit it was essential for the wires to be accurately located in relation to the cathode films and each other. For a spacing between the wires of 4 mm the tolerance was ± 10 μm. This accuracy was achieved by etching tracks in printed-circuit material, forming grooves in it to a depth of 70 μm and positioning the wires with the help of the grooves. The grooves were made using a numerically controlled milling machine at the Philips Engineering Workshops in Eindhoven. The machine was fitted with a slotting mechanism, with a diamond tool, designed for this application (*fig. 7*).

The wires were attached in position by 'reflow' soldering. In this process a tin layer applied earlier is made fluid again with an electrode that can be set for time, temperature and pressure. Besides creating a good soldered joint this method also ensured that the wire was located at the bottom of the groove (*fig. 8*). A special semi-automatic machine was built for positioning, tensioning and soldering the large numbers of wires (there were about 2400) [4].

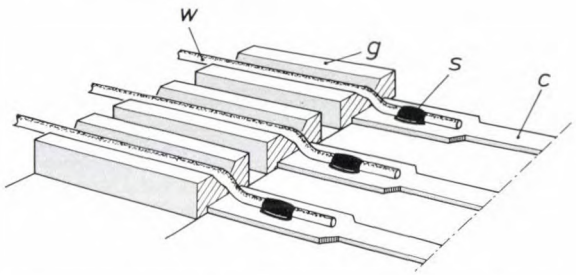
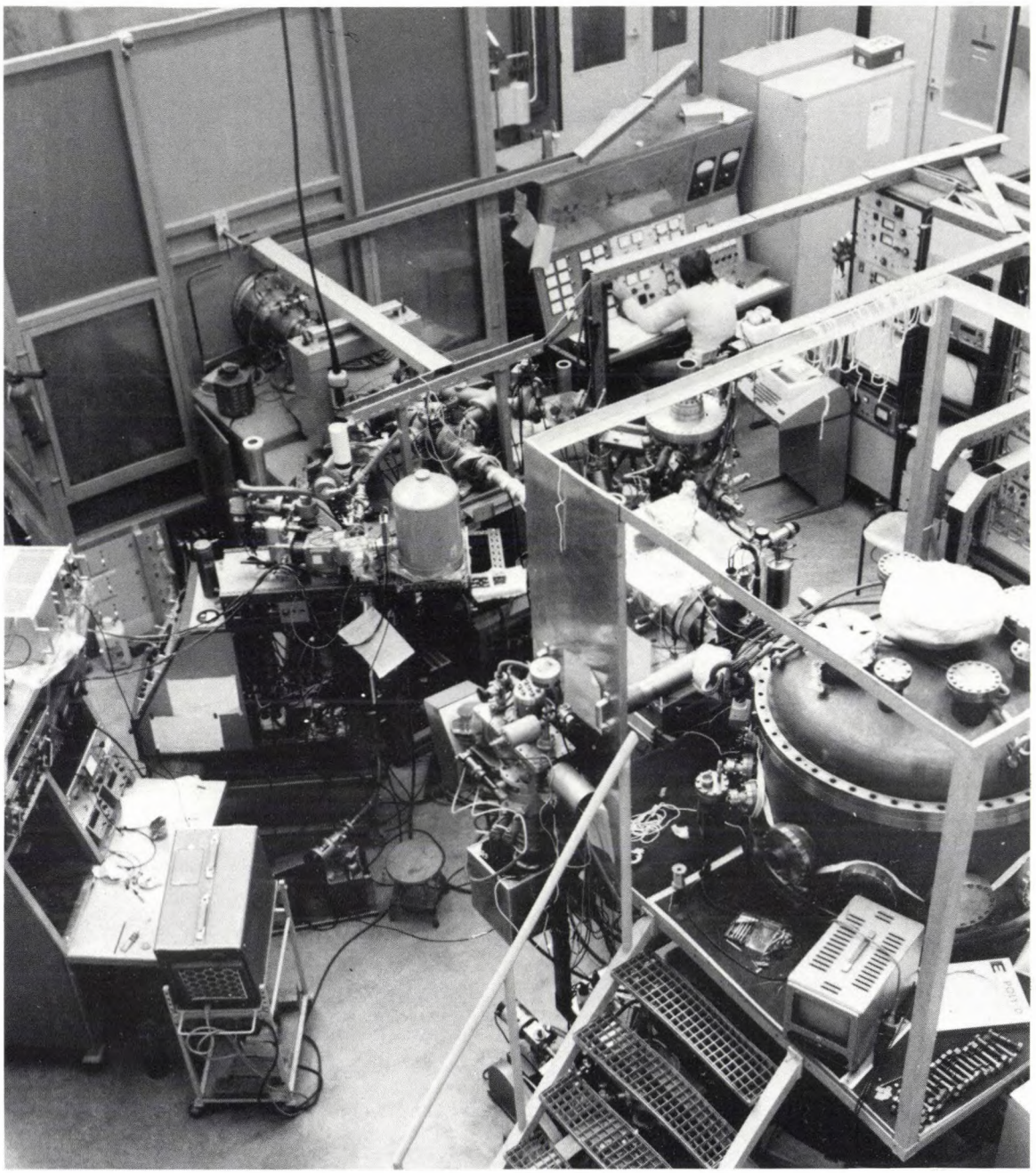


Fig. 8. Soldered joints *s* of tungsten wires *w* on copper contacts *c* in a 'wire chamber'; the position of the wires is determined by the grooved strip *g*.

[4] H. J. Akkerman, R. P. Arink and P. H. Thobe, Mikroniek **20**, 230, 1980 (in Dutch).



A 200 kV ion-implantation system in the FOM Institute for Atomic and Molecular Physics. This Institute is located at Watergraafsmeer, like IKO; the Philips group at IKO is working closely with the Institute on research into ion beams. The 200 kV source is located in the cage that can be seen in the upper left-hand corner. The ion beam generated here can be guided by means of a magnetic switch to any one of three experimental beam lines:
— a bench for ion scattering — Rutherford Back-Scattering — that was installed by Philips (the left-hand beam line, as seen from the high-voltage side),
— a bench for surface-structure investigations (the central beam line),
— a bench for measuring ion-induced X-ray radiation (the right-hand beam line).

Ion-beam technology

W. K. Hofker

One of the applications of ion-beam technology is in the manufacture of transistors, integrated circuits — ICs, 'chips' — and semiconductor detectors. In this application the ions from the beam are 'implanted' in the surface layer of semiconductor material [1]. An ion beam can also be used to analyse surface layers. An important application here is the analysis of implanted layers. The information obtained in this way can be used to improve the implantation process. Ion-beam technology is therefore important in manufacturing and research.

The Philips group at IKO started to do experiments on ion implantation in 1962. Silicon was doped with ions of phosphorus and lithium to make semiconductor detectors. Although this work on semiconductor detectors was a logical continuation of earlier work by the group [2] it was soon clear that ion implantation could be very useful in the manufacture of other electronic components, such as transistors, and the work was therefore expanded to cover such activities. Today IC technology is inconceivable without ion implantation: with other methods it is almost impossible to obtain the desired electrical characteristics to the same high accuracy. Consequently, there was an increasingly close cooperation with other IC groups within Philips, and this was one of the reasons why it was decided to transfer the Amsterdam group to Eindhoven. At present the group is working mainly on special implantation technologies, such as implantation at high ion energy and large dose; these new technologies require both fundamental and applied research. In this work it is also important to be able to obtain certain chemical properties, as well as electrical ones.

Experimental methods

As we noted earlier, ion beams are also used for investigating doped semiconductor material. The material can be bombarded with light ions, e.g. of hydrogen or helium. When they collide with atoms, there is back-scattering. Measuring the energy and number of the back-scattered ions at a given angle to the surface gives information about the composition and structure of the surface layer of the material (High-Energy Ion Scattering: HEIS; Rutherford Back-Scattering: RBS). It is also possible to bombard the material with ions and 'peel away' the surface layer by layer, i.e. by 'sputtering'; the ions released are then analysed with a mass spectrometer (Secondary-Ion Mass Spectrometry: SIMS). Finally, ion beams can be used to measure depth profiles by means of nuclear reactions.

In the earliest work on ion implantation the ions were supplied by a mass separator from IKO [3] (its acceleration voltage was 100 kV). *Fig. 1* shows a cut-away drawing of the machine. This equipment was later replaced by a machine for high-energy implantations (acceleration voltage 500 kV) and one for large-dose implantations (ion current 2 mA) [1]. HEIS analyses of implanted materials were performed with the aid of a Van de Graaff accelerator (3 MV) from the State University of Utrecht, in cooperation with the FOM (the Organization for Fundamental Research on Matter) Institute for Atomic and Molecular Physics in Amsterdam, where the necessary test equipment was already available. Since 1978 the group has been using its own instruments and equipment, which were housed at IKO until the group moved to Eindhoven.

Some experimental results

Depth distribution of boron ions

When ions are implanted in a material such as silicon, the crystal lattice is damaged, and this causes changes in the electrical behaviour. The damage can be repaired, however, by annealing. One of the things

Dr Ir W. K. Hofker is with Philips Research Laboratories, Amsterdam Department.

[1] W. K. Hofker and J. Politiek, Ion implantation in semiconductors, Philips tech. Rev. 39, 1-14, 1980. This contains a description of the implantation equipment and the HEIS and SIMS methods.
[2] See for example the articles by W. K. Hofker in this issue: Geiger-Müller counters, page 296, and Semiconductor detectors, page 298.
[3] K. J. van Oostrum and J. H. Dijkstra, Nucl. Instr. Meth. 29, 231, 1964.
[4] W. K. Hofker, H. W. Werner, D. P. Oosthoek and H. A. M. de Grefte, Radiation Eff. 17, 83, 1973.
[5] W. K. Hofker, H. W. Werner, D. P. Oosthoek and H. A. M. de Grefte, in: B. L. Crowder (ed.), Ion implantation in semiconductors and other materials, page 133; Plenum Press, New York 1974.
[6] W. K. Hofker, Philips Res. Repts Suppl. 1975, No. 8.
[7] W. K. Hofker, H. W. Werner, D. P. Oosthoek and N. J. Koeman, in: S. Namba (ed.), Ion implantation in semiconductors (Proc. 4th Int. Conf., Osaka 1974), page 201; Plenum Press, New York 1975.

we have investigated is the effect of such a treatment on the depth distribution of boron ions. It was for such studies that we, in cooperation with H. W. Werner and H. A. M. de Grefte of Philips Research Laboratories, Eindhoven, developed the SIMS method [4]. In this method the mass spectrometer

ditions for which the implantation and annealing will produce the desired electrical properties. SIMS is a test method much in use today and the analytical description has been taken up by many working in IC technology.

The depth distribution of boron ions was also

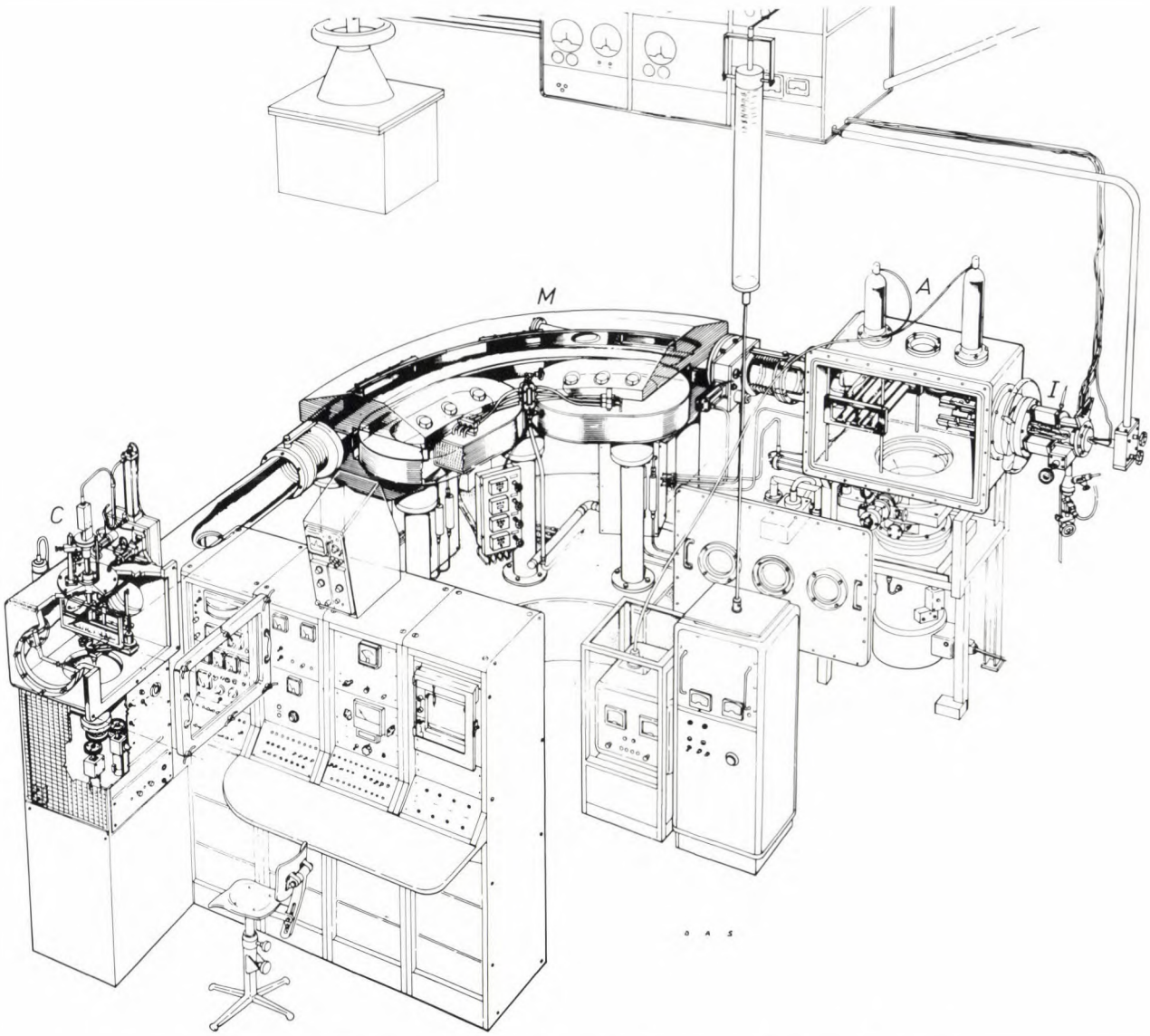


Fig. 1. Cutaway drawing of a mass separator from IKO that was used for ion implantation in an earlier phase of the work. *I* ion source, *A* 100 kV accelerator tube, *M* analysing magnet, *C* chamber in which the ions are collected.

is set to the boron line and the boron ion current is recorded as a function of time. The sputter rate is also measured. When the instrument has been calibrated, the boron concentration can be derived as a function of the depth from the measured results (*fig. 2*). We were also able to give an analytical description of the measured variation with depth [5][6]. From the experimental results it was possible to specify the con-

found to be affected by the implantation of other ions before the annealing [7]. We implanted various kinds of ions in silicon that had been homogeneously doped with boron (at a concentration of 1.3×10^{19} boron ions/cm^3). The doses were selected in such a way that the implanted areas in the silicon crystal became amorphous. After the annealing (at 800 °C) we found peaks in the originally uniform depth distribution of

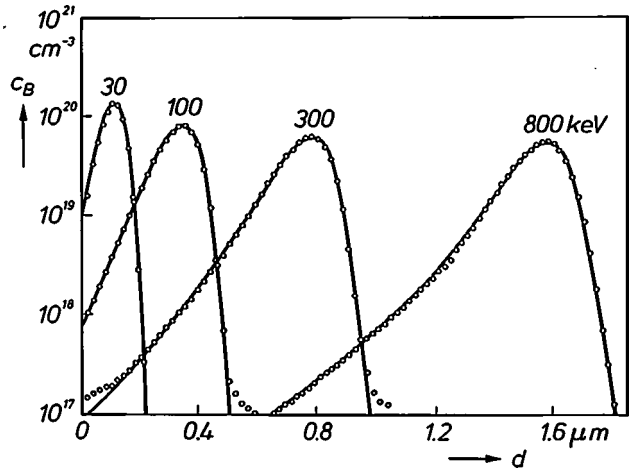


Fig. 2. Depth distributions determined by SIMS of boron (concentration c_B as a function of depth d) implanted in polycrystalline silicon with a dose of 10^{15} ions/cm^2 and energies of 30, 100, 300 and 800 keV. The points correspond to the measured values. The solid curves represent an analytical description of the measured results. (The analytical description was obtained [5][6] by determining the four statistical moments: mean value, variance, skewness and kurtosis for the measured points for each implantation energy and introducing these into a Pearson IV distribution.) The figure shows how good the agreement is. The depth profile for any one implantation energy can thus be calculated readily from a few measured profiles.

the boron ions (*fig. 3*). The location of the peaks coincided exactly with the points where the amorphous layer was in contact with the unchanged crystalline material. The redistribution of the boron that we have observed may therefore be due to the considerable local mechanical stresses [6].

In double implantations, which are often used in IC technology, the implantation profiles may also have an adverse effect on one another [8].

Damage caused by implantation

Even after annealing, some damage may still remain, usually in the form of dislocations in the crystal. These can be studied with an electron microscope.

Our research has shown that the occurrence of dislocations, e.g. after the implantation of antimony in silicon, is closely dependent on the temperature of the silicon during the implantation, even though it may be annealed later at a much higher temperature. For example, a dislocation network will be formed (*fig. 4*) if the temperature of the silicon during the antimony implantation is 150 °C and the silicon is annealed later (8 hours at 1100 °C). However, no dislocation network is formed if implantation at a temperature twenty degrees higher or lower is followed by a similar annealing procedure. Closer investigation has revealed that the network is due to lattice defects and mechanical stresses in the lattice, caused by the implanted ions. The combined effect of lattice defects and stresses is greatest at 150 °C.

Oxidation properties

Oxidation processes — which often take place at a temperature of 1000 °C — can play an important part in the manufacture of integrated circuits. It is well known that the presence of a thin layer of Si_3N_4 on the silicon inhibits the oxidation of the silicon. We recently carried out a study [9] to find out whether the oxidation processes can be regulated locally by implanting nitrogen in the silicon. A relatively small dose of nitrogen (3×10^{15} ions/cm^2) appeared to be sufficient to inhibit oxidation for an hour at 1000 °C.

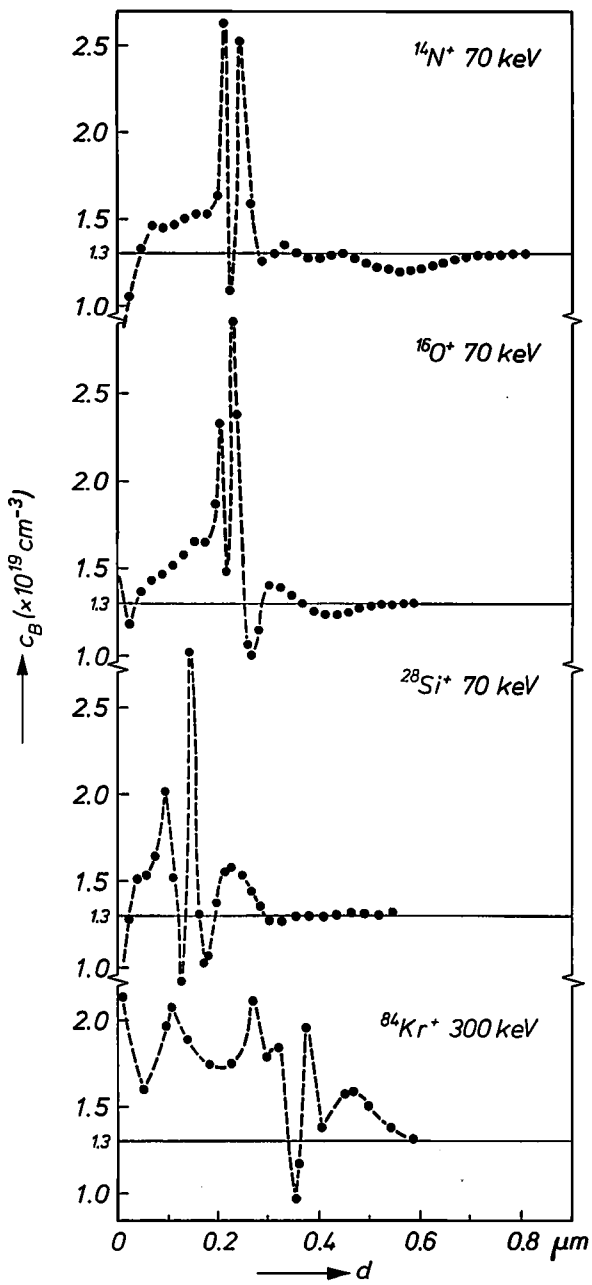


Fig. 3. Boron concentration c_B as a function of depth d after a silicon substrate, uniformly doped with boron, has been implanted with either nitrogen, oxygen, silicon or krypton, and has then been annealed for 40 minutes at 800 °C. The implantation energy for nitrogen, oxygen and silicon is 70 keV and for krypton it is 300 keV. The implantation dose is 10^{16} ions/cm^2.

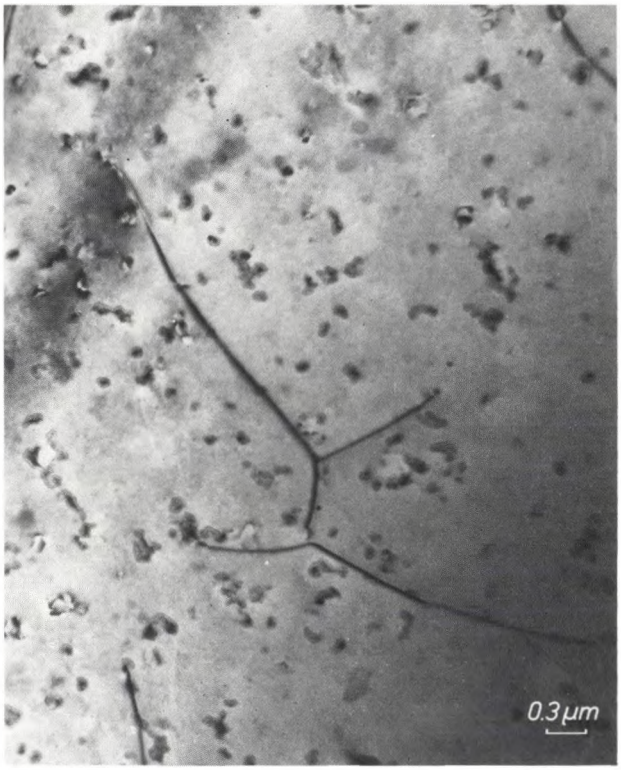


Fig. 4. A photomicrograph made with a transmission electron microscope of a silicon substrate implanted with antimony (energy 150 keV, dose 10^{16} ions/cm^2) and annealed for 8 hours at 1100 °C. During the implantation the temperature of the substrate was 150 °C.

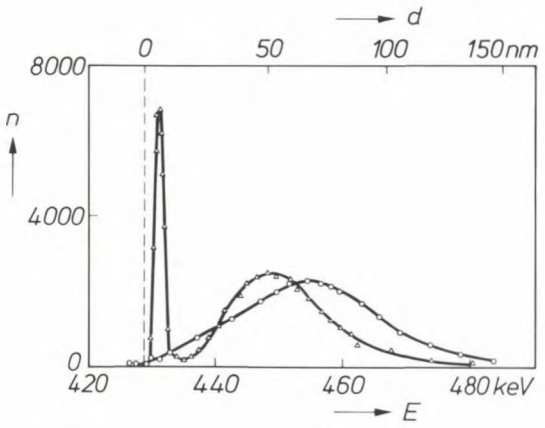
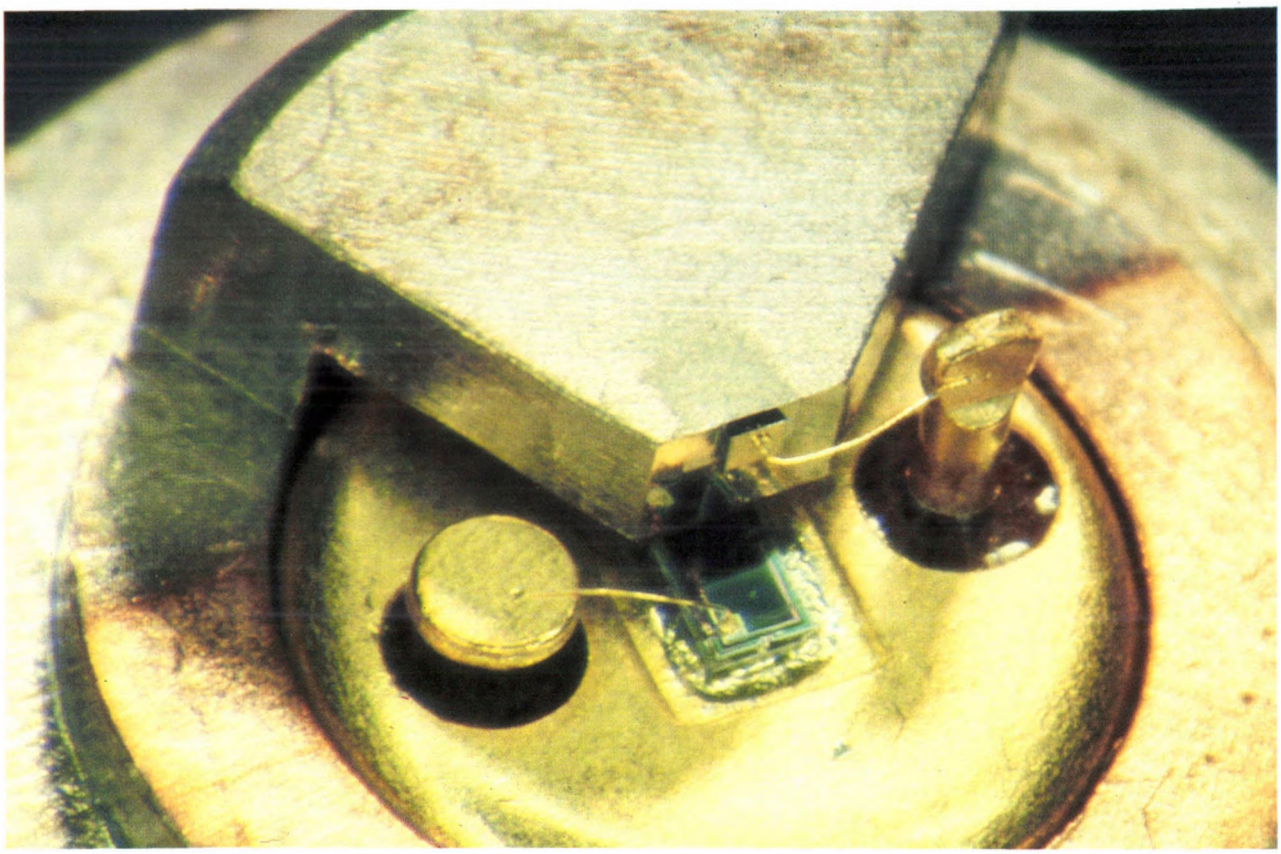


Fig. 5. Nitrogen distribution before (○) and after (△) annealing for an hour at 1000 °C in a silicon substrate implanted with nitrogen (dose 5×10^{15} ions/cm^2, energy 50 keV). The measurements were made with a resonant (p,$\alpha\gamma$) nuclear reaction. The horizontal axis gives the proton energy E and the vertical axis the number of gamma quanta n observed at this energy. The vertical dashed line indicates the position of the surface. The depth d is plotted horizontally across the top.

To obtain a clearer understanding of the inhibiting mechanism we analysed the depth distribution of the nitrogen, making use of a resonant nuclear reaction: if a ^{15}N nucleus is struck by a proton (hydrogen nucleus) of suitable energy, a carbon nucleus is formed, accompanied by the emission of a helium nucleus and gamma radiation; if the energy of the proton deviates only slightly this process will not take place. The incident protons lose energy when they pass through the material, so that the depth at which the protons have the correct energy for this nuclear reaction is associated with the energy with which the protons strike the material. We implanted ^{15}N in silicon, bombarded the material with protons and recorded the intensity of the resultant gamma radiation as a function of the energy of the incident protons (*fig. 5*). The figure shows that after annealing for an hour at 1000 °C nitrogen accumulated just below the surface. Further investigations showed that the accumulation of nitrogen at the surface is maintained during oxidation, while the nitrogen located deeper in the material gradually disappears. We assume that the inhibition of oxidation is due to formation of an Si_3N_4 monolayer at the surface. During the oxidation silicon-nitrogen bonds are broken and replaced by silicon-oxygen bonds. Because this is a slow process, only very little SiO_2 is formed. The Si_3N_4 layer is continuously replenished by diffusion of nitrogen from the supply located just below the surface — it is this that inhibits the oxidation — until the supply has been used up. Oxidation of the silicon will then proceed much more rapidly.

Oxidation can therefore be inhibited locally by implanting nitrogen at the desired locations, with the aid of a mask. This procedure will soon be used in the manufacture of CMOS integrated circuits (CMOS = Complementary Metal-Oxide Semiconductor) [10]. One of the electrodes (the 'gate') is oxidized — the oxide layer acts as an insulator — while the other electrodes (the 'source' and 'drain') do not become oxidized at the same time because of the implanted nitrogen. This technology can be used to produce a CMOS circuit of half the area of one made without implanted nitrogen: its potential is considerable.

[8] W. K. Hofker, H. W. Werner, D. P. Oosthoek and N. J. Koeman, in: G. Carter, J. S. Colligon and W. A. Grant (ed.), Applications of ion beams to materials 1975 (Inst. Phys. Conf. Ser. No. 28, 1976), page 13.
[9] W. J. M. J. Josquin and Y. Tamminga, unpublished.
[10] B. B. M. Brandt, W. Steinmaier and A. J. Strachan, Philips tech. Rev. **34**, 19, 1974.



Microscope photograph of a proton-bombarded double hetero-junction AlGaAs laser (CQL10) mounted on a copper heat sink. The dimensions of the crystal are 250 × 300 × 80 μm. The proton bombardment is made in such a way that the damage produced can be used to define the active region in the lateral direction. One advantage of this method is that by selecting the correct energy the damaged area can be very accurately located at the desired depth. The photodiode in the base stabilizes the output of the laser by feedback to the laser power supply.

Linear electron accelerators

P. Bakker

During 1980 the first test results were obtained from IKO's new electron accelerator, the Medium-Energy Accelerator (MEA) (*fig. 1*). It is planned to bring this accelerator into service in 1981 or 1982. The MEA has been designed to supply electrons at energies up to 500 MeV, with a maximum peak current of 20 mA, a peak width of 50 μs and a maximum repetition frequency of 2 kHz [1]. The relatively large duty cycle (maximum 10%) is particularly suitable for coincidence experiments, in which observations are made of the scattered electron and a nucleon released from the atomic nucleus after an electron has collided with an atomic nucleus. The equipment constructed for this purpose (*fig. 2*) has already provided excellent test results.

The primary electron beam will also produce secondary beams of pions and muons and these in turn can be used for investigating atomic nuclei.

The construction of the MEA and the nuclear instrumentation and equipment to go with it became a practical possibility when the Minister of Education and Science gave his approval to the plan in 1972 and made the necessary funds available — about ten million dollars at 1971 values. The Minister also decided that IKO should combine with the Dutch High-Energy Physicists to form a National Institute to be known as NIKHEF (National Institute for Nuclear and High-Energy Physics) and that IKO's two existing accelerators — the synchrocyclotron [2] and the EVA electron accelerator [3] — should be shut down; the accelerators were shut down in 1976 and 1977.

The story behind the MEA is quite a long one and one in which Philips also played an important part. In about 1957 Prof. P. C. Gugelot, who was then the Director of IKO, had the idea of building a linear accelerator for electrons with an energy of 3000 MeV, to complement the synchrocyclotron. However, the plan did not seem to be practicable, because of insufficiencies in money and manpower, and so it was decided, in consultation with Philips Research Laboratories, to build a much smaller machine (60 MeV). At Philips the main interest in building such an accelerator was the chance of acquiring greater experience in the specialized technologies of high vacuum, microwave techniques (klystrons), high-voltage work and precision engineering. The Philips activity would supplement the experience already gained at Mullard Research Laboratories at Redhill, England in the construction of low-energy accelerators (about 20 MeV) for medical use.

In 1961 a team of Philips employees led by Ir R. A. Koolhof started the development of the injector, modulator, waveguides, vacuum and cooling system, and the first experiments with the accelerator were made in 1964 for the radiochemical department of IKO. The accelerator was in use until 1967; it was then succeeded by EVA (*Elektronen Versneller Amsterdam*), a linear accelerator for electrons of energies up to 85 MeV.

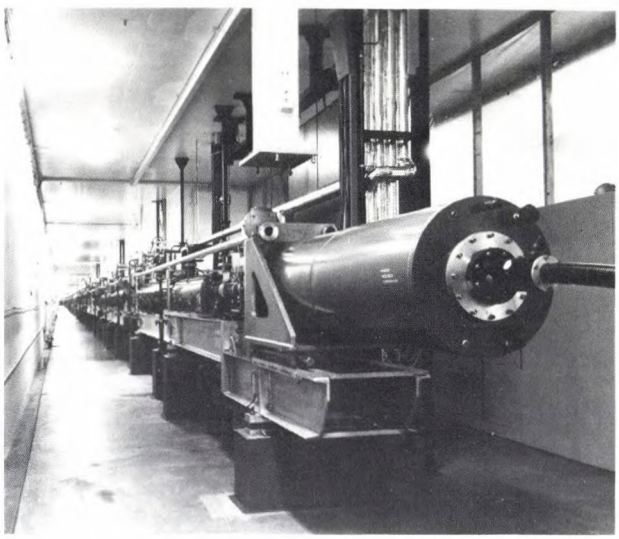


Fig. 1. The MEA, the Medium-Energy Accelerator, IKO's new two-hundred-metre long accelerator for electrons up to 500 MeV.

P. Bakker is with Philips Research Laboratories, Amsterdam Department.

This article arose partly out of discussions with Ir R. A. Koolhof (formerly with Philips Research Laboratories, Amsterdam Department, now with Philips Elcoma Division, Eindhoven) and Prof. Dr C. de Vries (IKO).

[1] C. de Vries and P. J. T. Bruinsma, Atoomenergie **18**, 98, 1976 (in Dutch).
[2] G. Luijckx, The cyclotron, this issue, page 290.
[3] P. J. T. Bruinsma, J. G. Noomen and C. de Vries, Nucl. Instr. Meth. **74**, 1, 1969.

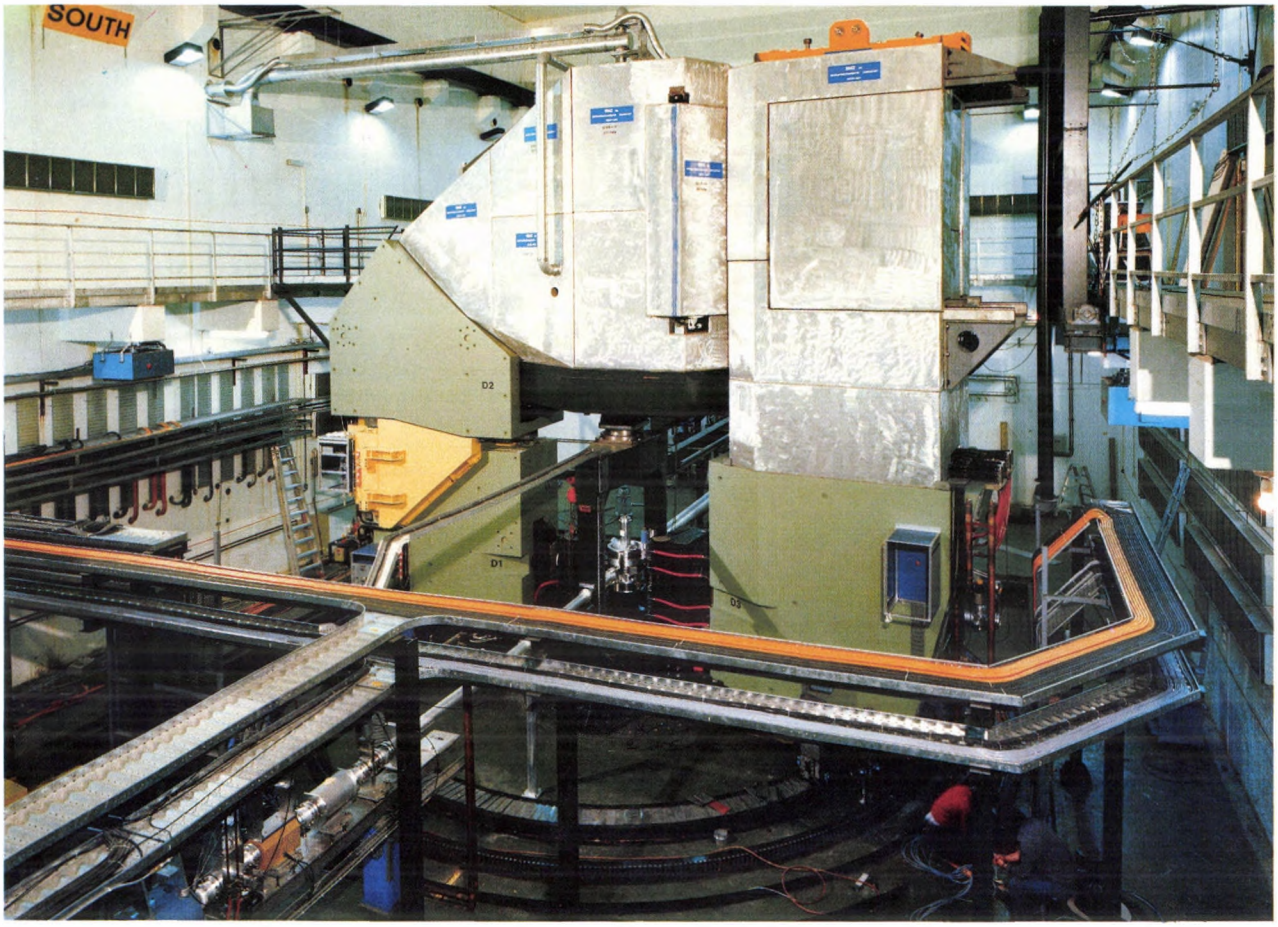


Fig. 2. Equipment for nuclear-physics experiments with the MEA. The spectrometer for observing electrons can be seen on the left, and the spectrometer for observing heavier particles released when an accelerated electron collides with an atomic nucleus in a target can be seen on the right. The two spectrometers can be rotated about the target. (The total height of the spectrometers is about 12 m.)

Principle of a linear electron accelerator

In a linear electron accelerator an electromagnetic wave travels along a waveguide. The field has an electric component in the direction of propagation of the wave. The electromagnetic power is obtained from a klystron driven by a pulsed high-voltage modulator, so that the power enters the waveguide in pulses. Electrons are injected simultaneously with the wave pulse. The electrons are accelerated by the electric component of the field. The phase velocity of the travelling electromagnetic wave is made equal to the velocity of the electrons (almost equal to the velocity of light). The acceleration process is maintained in this way over the entire length of the waveguide. The correct phase velocity is obtained by placing diaphragms in the waveguide.

In the construction of the accelerator the finish of the interior surfaces of the waveguide structure must be of the highest quality and the dimensional accuracy must be exceptionally good. The injection of the electrons also requires the greatest care, particularly if the spread in energy of the accelerated electrons is required to be small.

The Philips accelerator

Something has already been said about the earlier history of the Philips accelerator. The design was for an accelerator for 60 MeV with an energy spread of 1%, a mean beam current of 200 μA, a repetition frequency of 200 Hz and a duty cycle of 0.1%. Most of the components (*fig. 3*) were developed by Philips. The manufacture of the waveguides, for example, was in general terms as follows.

Copper discs and aluminium rings were machined on a precision lathe with hydrostatic bearings [4]. The discs and rings were then stacked alternatively to form a 'pipe' about a metre long and were then clamped

together (*fig. 4*). A copper layer about 1 cm thick was electroplated on to the outside of this pipe. The aluminium was next removed by selective etching. After a brief finishing treatment the waveguide was ready.

Further development

In the early sixties the use of electron accelerators began to increase rapidly on an international scale and in about 1965 the requirements for the quality (energy spread, duty cycle) of the beam began to become more exacting, especially if it was to be used for nuclear-physics research. Further developments in this direction, however, no longer came within the purview of Philips Research Laboratories. IKO were fortunate in that they were able to obtain two accelerator sections on loan from the Stanford Linear Accelerator Center (SLAC) in California. These were manufactured for a two-mile long electron accelerator that had been constructed at SLAC, containing about a thousand sections each three metres long. A modulator to drive these accelerator sections was built at IKO. Substantial help was given here by the Philips people who had worked on the construction of Philips' own accelerator. In 1967 EVA was used for the first nuclear-physics experiments. The research carried out with this accelerator generated some twenty doctoral theses, ten dissertations and twenty-five publications.

In the second half of the sixties international interest in research with electrons moved more towards precision and coincidence measurements at higher energies. These developments and the experience that IKO had gained with electron accelerators eventually led to the construction of the MEA.

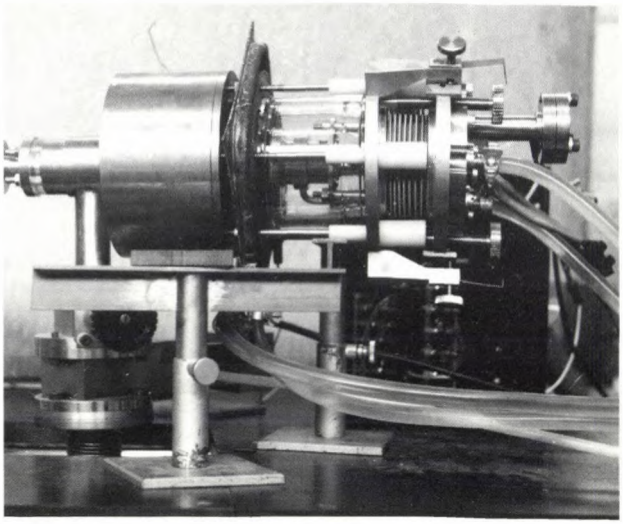


Fig. 3. The injector developed by Philips for the Philips accelerator, which was in use until 1967 at IKO, mainly for radiochemical experiments.

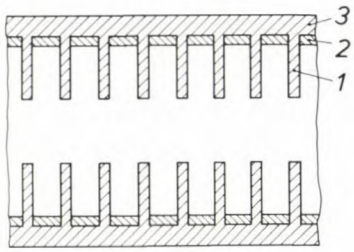


Fig. 4. Sketch of the corrugated waveguide of the Philips accelerator. Accurately machined copper discs (*1*) and aluminium rings (*2*) were stacked alternately to form a pipe and then clamped together. A copper coating (*3*) was electroplated on to the outside. The aluminium rings were then removed by selective etching.

[4] H. J. J. Kraakman and J. G. C. de Gast, A precision lathe with hydrostatic bearings and drive, Philips tech. Rev. **30**, 117-133, 1969.

Copyright: Aerophoto-Schiphol b.v.

Recent aerial photograph of the Scientific Centre at Watergraafsmeer in the Netherlands; as well as the National Institute for Nuclear and High-Energy Physics the site also contains the FOM Institute for Atomic and Molecular Physics, the Amsterdam Academic Computer Centre, the 'Anna's Hoeve' Biological Centre of the University of Amsterdam and the Mathematics Centre.

Volume 39, 1980, No. 11 pages 285-328 Published 20th November 1981

Supplement to Philips Technical Review
Beilage der Philips Technischen Rundschau
Bijlage van Philips Technisch Tijdschrift

PHILIPS

November 1981

Recent United States Patents

Abstracts from patents that describe inventions from the following research laboratories that form part of or cooperate with the Philips group of companies:

Philips Research Laboratories, Eindhoven, The Netherlands	*E*
Philips Research Laboratories, Redhill, Surrey RH1 5HA, England	*R*
Laboratoires d'Electronique et de Physique Appliquée, 3 avenue Descartes, 94450 Limeil-Brévannes, France	*L*
Philips GmbH Forschungslaboratorium Aachen, Weißhausstraße, 51 Aachen, Germany	*A*
Philips GmbH Forschungslaboratorium Hamburg, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany	*H*
Philips Research Laboratory Brussels, 2 avenue Van Becelaere, 1170 Brussels (Boitsfort), Belgium	*B*
Philips Laboratories, N.A.P.C., 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, USA	*N*

4 253 880

Device for the conversion of solar energy into electrical energy

J. Bellugue *L*

A device is described for the conversion of solar energy into electrical energy with the aid of a photo-voltaic cell. A central lens forms a round radiation spot on the radiation-sensitive area of the cell, while a toric mirror arranged round the lens forms an annular radiation spot. Thus, it is ensured that the entire area of the cell remains illuminated, independent of the movement of the sun relative to the cell. Owing to the selected intensity distribution in the radiation spots a satisfactory efficiency of the cell can be maintained even at a high concentration of the solar energy.

---

4 254 427

Semiconductor device having a compact read-only memory

J. Lohstroh *E*

A read-only memory in which each memory cell is formed by two back-to-back diodes across which a connection can be formed by means of punch-through. Since cross-talk between adjacent cells is impossible, the packing density may be very large. Additionally, the cycle time of the memory is low due to the very short reverse recovery time of the invented structure.

---

4 254 820

Heat transport device

G. A. A. Asselman *E*
J. Schröder
F. Mahdjuri

There is provided heat transport apparatus comprising a closed tube having an evaporator and a condenser with the condenser being positioned above the evaporator during operation. The closed tube includes a heat transport medium which, during operation, is vaporized in the evaporator and condensed in the condenser, the resulting condensate being returned to the evaporator. A valve is positioned in the closed tube in the condenser region for controlling the condensate return. A pressure-expansion chamber is arranged within the closed tube and is connected to the valve for closing the same when the vapor pressure in the condenser exceeds a given value. A pressure-equalization duct forms part of the valve and provides communication between the evaporator and the condenser.

4 254 442

Circuit for the protection of telephone lines

E. C. Dijkmans *E*
F. A. C. M. Schoofs

Arrangement for protecting telephone lines and electronic circuits connected thereto from overvoltages. A first pair of diodes clamp the voltage of each wire versus the supply voltages in cooperation with a current limiting impedance. A second pair of (Zener-)diodes, separated from the first diode pair by means of a current limiting impedance, clamp the forward 'transient' produced across the first diode pair versus the same or slightly different supply voltages.

---

4 255 677

Charge pump substrate bias generator

L. Boonstra *E*
C. W. Lambrechtse
R. H. W. Salters
R. M. G. Wijnhoven

A semiconductor device having an integrated circuit of which a region of one conductivity type is charged by supplying charge carriers from a zone of the opposite conductivity type to an inversion layer formed in the said region below a field electrode at which a voltage is set up. When the voltage is switched off, a part of the charge carriers recombine in the said region. According to the invention, the charge carriers are supplied from a supply conductor and an electronic switch is present between said conductor and the inversion layer, which switch prevents the flow back of charge carriers to the supply conductor when the inversion layer disappears.

---

4 257 854

Method of producing objects with a supersmooth aluminum surface

T. E. G. Daenen *E*
J. M. Oomen
J. F. M. van de Berg

Method of producing objects having an aluminum surface of a supersmooth quality by applying an aluminum layer which is at least 10 micron thick onto the surface of the object from an electroplating bath and by thereafter subjecting the object to a mechanical machining precision operation. The aluminum layer is preferably electroplated by means of an anhydrous bath with an aprotic solvent.

4 258 114

Electrophotographic recording material and method of manufacturing same

R. Clasen *A*
H. G. Junginger

An electrophotographic recording material having a porous layer of photoconductor-binder, particularly with crystalline tetragonal lead monoxide as the photoconductor, is provided between an electrically conductive layer and a dielectric foil, said pores of the layer of photoconductor-binder, prior to providing the foil, being filled with a high-ohmic dielectric liquid to wet both the layer of photoconductor-binder and the foil. Preferably, the pores of the layer are filled with tetramethyl tin as the photoconductor-binder. As a result of this, the use of an adhesive between the foil and the layer of photoconductor-binder may be omitted. A large porosity and hence a great sensitivity of the layer of photoconductor-binder are maintained.

---

4 258 256

Device for examining a body by means of penetrating radiation

G. Harding *H*

A body is irradiated by a primary radiation beam; the scattered radiation produced thereby is incident on a detector device consisting of a plurality of detectors and is a measure for the density distribution of materials in the primary beam. However, this gives rise to multiple scattered radiation which disturbs the measurement. The present invention provides a device in which disturbance by multiple scattered radiation is greatly reduced, without any significant effect on the scattered radiation originating from the primary beam. A device in accordance with the invention is provided with flat laminations which are arranged between the primary beam and the detector device and which are aligned with the primary beam in a fan of flat planes.

---

4 258 282

Device for the generation of a control voltage across a piezo-electric positioning element

A. M. A. Rijckaert *E*

A device for the generation of a control voltage across a piezo-electric positioning element. In order to enable a high voltage to be generated across the piezo-electric element without using a comparatively high supply voltage, use is made of the capacitance of such an element as an integrating element. The device comprises a current source and switches for transferring current pulses to and from the piezo-electric element on command of control pulses.

---

4 258 554

Refrigerator

G. A. A. Asselman *E*
A. J. van Mensvoort

A refrigerator comprising a freezing compartment and a refrigerating compartment which refrigerator is provided with a primary refrigerating system containing a refrigerant with a primary evaporator disposed in the freezing compartment, and a secondary refrigerating system which also contains a refrigerant with a secondary evaporator disposed in the refrigerating compartment, and a secondary condensor which is in heat-exchanging contact with the primary evaporator, which condensor has a condensation wall on whose surface the refrigerant condenses during operation, means being provided for varying the available condensation wall area, so as to control the temperature of the secondary evaporator. Preferably, the secondary condensor is provided with a reservoir containing a control gas, which during operation constitutes an interface with the refrigerant vapor at the location of the condensation wall with the aid of a reversible control gas getter, which can be heated and which is located in the reservoir, which getter enables the amount of free control gas to be varied. The reversible control-gas getter can be heated by means of an electric heating element which is included in an electrical control circuit.

---

4 259 435

Additive method of manufacturing metal patterns on synthetic resin substrates

D. J. Broer *E*
A. J. M. van den Broek

Additive method of manufacturing metal patterns on synthetic resin substrates, printed circuits in particular. A nuclear image of a desired pattern is obtained by means of a salt of a metal which may occur in more than one valency and a noble metal salt by inactivating the substrate area outside the desired pattern by photo-oxidation by means of exposure to light in the presence of oxygen. The ultimate metal pattern is deposited by electroless metal deposition on the nuclear image.

---

4 259 678

Semiconductor device and method of manufacturing same, as well as a pick-up device and a display device having such a semiconductor device

G. G. P. van Gorkom *E*
A. M. E. Hoeberechts

The invention relates to a semiconductor cathode based on avalanche breakdown in the *p-n* junction. The released electrons obtain extra accelerating energy by means of an electrode provided on the device. The achieved efficiency increase makes the manufacture of such cathodes in planar silicon technology sensible. Such cathodes are applied, for example, in cathode ray tubes, flat displays, pick-up tubes and electron lithography.

---

4 260 663

Method of producing a dielectric having a perowskite structure

H. J. Hagemann *A*
S. Hunten

Only those dielectrics which have a correspondingly low sintering temperature are suitable for the production of ceramic multi-layer capacitors having intermediate electrodes consisting of non-noble metals, such as nickel and cobalt. Suitable are ceramic masses comprising alkaline earth zirconates, wherein up to 10 mole% of the zirconium may have been replaced by titanium; iron or nickel of manganese being added as doping agents to the stoichiometrical perowskite basic compound having a composition of alkaline earth metal $(Zr_{1-x}Ti_x)O_3$ where $O < x \leqq 0.07$ and the whole assembly being sintered in a reducing atmosphere.

---

4 260 887

Electroradiographic recording device

H. Dannert *A*
H.-J. Hirsch

An electroradiographic display with a high picture quality at low radiation intensities is obtained with a device consisting of a source of X-rays, an electrode passing X-rays, an intermediate recording space for an object to be displayed, an electrode on the side remote from the recording space on which a layer of dielectric material is disposed, and a second electrode with a photoconductive layer. The layers are separated by a gas gap which is bounded by one or more side walls. A direct voltage source is in electrical contact with the electrodes. The photoconductive layer comprises a granular photoconductive material in a binder. The gas gap between the dielectric layer and the photoconductive layer is from 50 to 500 μm wide and the electrode passing X-rays has a surface resistance between 10^3 to 10^8 Ohms.

4 260 893

Device for directing electrically charged particles towards a target

P. Bakker E
R. S. Kuit
J. Politiek

A device for directing electrically charged particles towards a target present on a connection place of a support comprises an electrostatic deflection system having deflection plates for controlling a beam of charged particles in two mutually perpendicular direction towards the target with the deflection plates being brought at a desired voltage by means of an electronic control system, and two juxtaposed rods of electrically conductive material being provided on oppositely located sides of the target with the rods being each connected by a resistor to a fixed potential and being furthermore connected to a control device which reacts to a voltage which arises across the resistors if the beam of electrically charged particles impinges on a rod and which, if the beam, during each stroke across the target, impinges on a combination of rods other than the two innermost rods, produces a correction voltage for the control of the deflection plates.

---

4 260 897

Method of and device for implanting ions in a target

P. Bakker E
R. S. Kuit
J. Politiek

In a method of implanting ions in a target, an ion beam is directed onto the target by means of an electrostatic deflection system where the beam describes a pattern over the target dependent on voltage variations on two mutually perpendicular sets of deflection plates. One of the sets of deflection plates is subjected to a varying voltage difference and the other set of deflection plates is subjected to a constant voltage difference so that the ion beam describes a straight line on the target, while at the end of the beam stroke a fixed voltage difference is superimposed on the plates having a constant voltage difference and the variation in the voltage difference is reversed in the plates having a varying voltage difference in such manner that lines described consecutively by the ion beam are always parallel and are situated at a fixed distance from each other.

---

4 260 931

Low-pressure mercury vapor discharge lamp with luminescent coatings on envelope walls

G. A. Wesselink E
H. Roelofs
C. H. M. van Bommel

Low-pressure mercury vapor discharge lamps of the fluorescent type are known which comprise two nesting glass members so grooved as to provide between their nesting surfaces a circuitous ('folded') discharge path between two electrodes, the length of the discharge path being considerably greater than the physical distance between the electrodes. The whole surface of each glass member is coated with luminescent material with the result that, when the lamp is in use, the portions of the members not surrounding the discharge appear dark due to their poor light transmissibility. According to the invention, these portions of the two members are not coated with luminescent material and thus act as windows for light generated in the luminescent layers facing the lamp axis. Preferably, a fluted reflector is provided inside the lamp.

---

4 263 920

Method of and device for determining internal body structure

M. Tasto H
H. Schomberg

Determining the internal structure of a body, for example, a human body, by means of an electric field which extends between individual electrodes of an electrode array which at least partly surrounds the body. Each time the value of the currents flowing through the individual electrodes is measured. From the currents electrical resistance values in individual tubes of flux generated between the electrodes are measured. By exposure of the body successively to electric fields which each have a different direction and by measurement of the resistance values then occurring, the specific resistance in individual elements of a matrix which is imagined to be stationary with respect to the body can be determined. From the given distribution of the specific resistance in the elements of the matrix, resistance values are calculated in all tubes of flux successively at least once for each field direction. For determination of a correction for the specific resistance in the elements of the matrix, the calculated resistance values are compared with the measured resistance values. The magnitude of the difference determines the magnitude of the correction.

---

4 264 027

Capstan drive for a tape

H. Althuber H
H. Dimigen
H. Hübsch

The driving arrangement for a magnetic tape is comprised of a drive shaft having a smooth surface and a pressure roller having a high surface friction coefficient. The circumference of the drive shaft is provided with a thin layer of the borides, nitrides or carbides of a hard metal, which increases the friction resistance of the shaft. This measure reduces the slip, without affecting the proper uniform travel of the tape.

---

4 264 149

Liquid crystal display device

M. de Zwart E
T. W. Lathouwers

In a liquid crystal matrix display device in which the bistability effect of a cholesteric liquid crystal is used, the number of lines to be driven is increased by providing the electrodes on grooves provided in the supporting plates. The electrodes partly overlap the part of a supporting plate present between two grooves. The electrode surface and the remaining surface of the supporting plates are covered by a dielectric layer. The grooves have a depth which is equal to half the desired thickness of the liquid crystal layer. Herewith a uniform thickness of the liquid crystal layer is also obtained.

---

4 264 881

Microwave device provided with a half lambda resonator

F. C. de Ronde E

A microwave device is disclosed comprising a microstrip line pattern including an open ring forming a $\frac{1}{2}\lambda$ resonator having a narrow gap in which the electromagnetic field is closely tied to the ring.

---

4 264 975

Receiver for frequency modulated signals having two quadrature channels

J. O. Voorman E

A receiver for frequency modulated signals comprises two phase quadrature channels with synchronous demodulation and low-pass filtering. The object aimed at is to provide a tunable receiver without input filter. The absence of an input filter excludes the use of a limiter and the receiver will then be sensitive to amplitude variations. This problem has been solved by dividing the signals in the receiver in one or two stages by an amplitude factor which is derived from the output signals of the low-pass filters.

4 265 511

Detachable connector for optical fibres

A. J. A. Nicia E
C. J. T. Potters

A detachable connector, for coupling the ends of optical fibers, comprising two connector portions, each connector portion having fixing means, for detachably connecting the connector portions to each other, and at least one rotationally symmetrical housing having a bore which is coaxial to a central axis. The housing has a reference face at one end and adjusting means secured in the bore at an opposite end. The adjusting means enables an optical axis, of an end of an optical fiber to be secured in a tube, and to be adjusted parallel to the central axis of the housing. When the two interconnected connector portions contact each other by way of the reference faces, the optical axes of the fibers in each connector portion are parallel to one another.

---

4 266 146

Charge transfer devices having switchable blocking electrodes

A. H. M. van Roermond E

A charge transfer device having one or more control electrodes connected to switching means so that the electrode can perform two functions. Normally the electrode is used for charge transfer but it may be switched to act as a blocking electrode for blocking the charge present underneath it for one or more clock phases. This permits the device to be formed as a filter, an analog-to-digital converter or digital-to-analog converter in a simple and inexpensive manner. The charge transfer channel may be split into sections using isolation diffusions.

---

4 267 408

Arrangement for applying a signal to a transmission line

F. A. C. M. Schoofs E

An electronic circuit for applying a signal to a telephone line is disclosed wherein the signal to be transferred is superimposed on a direct current and applied to the reference voltage circuit of a bidirectional current source. The current source in turn applies, from high internal resistances, oppositely directed currents which are equal to each other and proportional to the signal onto the two wires of the telephone line.

---

4 268 777

Cathode-ray tube

J. H. T. van Roosmalen E

A cathode-ray tube having an electron gun to generate an electron beam and a focusing lens to focus the electron beam on a target. The anode of the electron gun forms part of the focusing lens and has a very small aperture to limit the electron beam. In order to prevent positive ions formed in the tube from poisoning the cathode the potential of the anode is at most 75 volts relative to the cathode potential and the distance from the center of the focusing lens to the aperture in the anode is at least equal to 1.5 times the largest dimension of the anode taken in a cross-section at right angles to the axis.

---

4 269 625

Bath for electroless depositing tin on substrates

A. Molenaar E

Method of currentless deposition of tin on a catalytic surface by means of a highly alkaline solution which contains stannous ions in a quantity of at least 0.20 mole/l and is used at temperatures of 60 to 90 °C. The solution operates on the basis of the mechanism of disproportioning of stannous ions. The tin deposition is, however, accelerated by means of a strong reducing agent such as a hypophosphite.

4 269 898

Resistance material

A. H. Boonstra E
C. A. H. A. Mutsaers
F. N. G. R. van der Kruijs

Resistance material consisting of a mixture of metal oxidic compounds, metal oxides, a permanent and a temporary binder, the resistance-determining component consisting of thallium rhodate $TlRh_2O_4$. This component has a linear positive temperature coefficient of the resistance *TCR* which enables the composition of a resistor having a very low *TCR* by combining it with a material having a negative *TCR*. The resistor is produced on the basis of this resistance material provided on a substrate.

---

4 270 045

Apparatus for reading an optical radiation-reflecting information carrier for controlling focus

W. G. Opheij E
P. F. Greve

An apparatus as described for reading an optical radiation-reflecting information carrier, which apparatus comprises an opto-electronic focusing error detection system with two focusing detectors. The path of the read beam which has been reflected by the information carrier and which originates from the read spot includes a single radiation-deflecting element, in such a way that the radiation which is incident on this element is deflected towards the focusing detectors.

---

4 270 960

Method of manufacturing a semiconductor device utilizing a mono-polycrystalline deposition on a predeposited amorphous layer

L. J. M. Bollen E
J. Goorissen

A method of manufacturing a semiconductor device is provided in which a masking layer is formed on a part of a surface of a monocrystalline semiconductor body and the semiconductor body is then subjected at the side of the surface to an epitaxial treatment from a gaseous phase, and an epitaxial layer is deposited of which a portion on the uncovered part of the surface is monocrystalline and a portion on the masking layer is polycrystalline. This method is characterized in that, prior to the epitaxial treatment, an amorphous or polycrystalline layer is deposited both on the masking layer and on the uncovered part of the surface at a temperature which is lower than that at which the epitaxial layer is deposited. In this layer the layer portion on the uncovered surface part changes into the monocrystalline state by a thermal treatment preceding the deposition of the epitaxial layer.

---

4 271 713

Speed- or movement-reducing drive

M. J. J. Dona E

A reduction drive having a drivable first structural member which presses a number of bodies of revolution against at least one thin intermediate plate. The intermediate plate is connected to the output of the drive and is supported by a second structural member. The material of the intermediate plate has a modulus of elasticity at least equal to that of the second structural member, and the bodies of revolution are pressed with sufficient force to locally but elastically deform the plate and second member.

---

4 272 389

Absorption heat pump work medium consisting of a solution of a fluorochloroalkane in a substituted amino phosphorus oxide

U. K. P. Bierman E

A work medium for use in an absorption heat pump consists of a solution of a chlorofluoroalkane in a substituted amino phosphorus oxide. With such a work medium it is possible for an absorption heat pump to absorb heat at − 20 °C and to give up heat at 70 °C.

PHILIPS TECHNICAL REVIEW

VOLUME 39, 1980, No. 12

Megadoc, a modular system for electronic document handling

J. A. de Vos

The ever-growing flood of paper is a sign of our times. In offices and in industry more and more valuable time is spent in filing documents — and trying to find them again. However, scientists at the Geldrop Project Centre of Philips Research Laboratories have now designed a system in which documents are dealt with by electronic methods alone. This is a modular system, which converts the information in each document into a binary-digital electronic signal that can then be processed, filed, retrieved, sorted, distributed or duplicated. A great stimulus to the development of the system was the appearance of a new medium for very compact storage of digital information — the DOR disc.

Introduction

In our daily life we now have to cope with a deepening flood of papers. At Philips in Eindhoven, for example, an average of one document of one or more pages is generated every day for every four employees and sent out by the public postal service. Similarly, an average of one document for every fourth employee arrives every day by external post and is then processed. This processing of documents consists in reading and classifying them and either storing or destroying them, straight away or after some time. In the day-to-day work it is necessary to consult these stored documents frequently, so that a great deal of time is spent searching in files of papers, sometimes unsuccessfully.

The search for methods for the electronic handling of documents has been going on for a long time [1]. Considerable progress has been made in some areas. A distinction should be drawn here between methods in which the document is handled pictorially, and those in which the document is processed alphanumerically, i.e. each character of the text is 'translated' into a binary-digital word. The first group of methods includes facsimile equipment, which can transmit documents over a telephone line. Because of the limited bandwidth of the telephone link it takes a long time, usually several minutes, to transmit a single document; a single document here and in the rest of this article will always be taken to be a single A4 page, unless otherwise indicated. An example of the second group of methods is the well-known telex equipment that seems such an essential feature of life today. Word processors also operate on the second principle. The use of such equipment, although only recently introduced, is increasing rapidly. Another approach that we should mention for the mass storage of documents is microfilm. Although microfilm has little or nothing to do with electronics, it can be used for storing large quantities of documents in a very small space.

The electronic processing of information in the methods mentioned above concentrates mainly on one aspect of the handling of documents, usually transporting them. Also, the documents have to be handled either pictorially or alphanumerically. The special feature of the Megadoc system developed by Philips Research Laboratories is that it deals with all aspects of the handling of documents: conversion into a digital electronic signal, classification, filing, retrieval, sorting, distribution and duplication. The document can also be processed both alphanumerically and pictorially, as desired. In principle, the document can also be entered into the system with information about grey levels or colours, although this is not yet possible with our system.

Dr J. A. de Vos is with Philips Research Laboratories, Eindhoven. He has led the team working on the Megadoc project at the Geldrop Project Centre for several years.

[1] See for example R. P. Uhlig, D. J. Farber and J. H. Bair, The office of the future, North-Holland, Amsterdam 1979, which contains many references to the literature.

Estimates of the total investment to the present day per office employee in the United States of America have given figures that are only a twentieth of the investment per agricultural employee and a tenth of the investment per industrial employee. In fact, there has been little change in the working methods and organization of office work through the years. The Megadoc system, however, could drastically alter the appearance of an office and what goes on in it. The 'document flow' in this office of the future might well take the following form:

The secretary opens the mail and uses the document reader to record the documents in the mass memory of the Megadoc system. Then she uses her computer terminal to tell the system to deliver the mail to the 'electronic mail boxes' of the people for whom she provides secretarial help. Each of these people can then use his own terminal, at any moment he chooses, to tell the system to display the documents addressed to him on his own document display. The system then asks the user — through his terminal — what should be done with each document: is it to be retained in the system or not? If the answer is yes, the system asks for some data and keywords to provide a document description that can then be stored in the data base. If at some later time it is necessary to call up a document from the system and to display it, then the user enters the data or keywords that he remembers. The system then presents, almost immediately, the descriptions of documents that match the criteria given. After the user has selected a document the system traces and reads the correct DOR disc within five seconds and the document appears on the document display. If he wishes, the user can also obtain a 'hard copy' (i.e. one on paper).

In the following sections of the article, the methods for 'converting' the document into the digital information required by the system will be described first of all. A brief account is then given of the way in which this information is filed in such a space-saving manner. This is followed by a description of the full system; more detailed attention is then given to certain modules that we have specially developed. Finally, the software that manages and controls the information flow is discussed.

Conversion of documents into digital information

The Megadoc system has been designed for processing documents of the European A4 format (210 × 297 mm) and the American A format (216 × 279 mm). The maximum area that can be handled is therefore 216 × 297 mm. The recommendations of the CCITT (Comité Consultatif International Télégraphique et Téléphonique) specify either the HR mode (high resolution) or the SR mode (standard resolution). In both modes the document is scanned in a horizontal direction with a resolution of 8 dots/mm; the resolution for the scan in the vertical direction is 7.7 lines/mm (see *fig. 1*) for the HR mode and 3.85 lines/mm for the SR mode. In the HR mode the document is thus subdivided into $297 \times 7.7 = 2287$ lines of $216 \times 8 = 1728$ picture elements or pixels, i.e. a total of 3 951 936 or approximately 4×10^6 pixels, see *fig. 2*. The specification for the Megadoc system quotes a time of one second for scan or display of a document, a value chosen because it is convenient for the user. To achieve this, no fewer than 4×10^6 pixels per second must be processed by the system.

In the line-by-line optical scan of the document the grey level is measured for each pixel. The range of possible grey levels has been reduced to just two levels in the present system: one above a given threshold level and one below it. Each pixel is therefore designated as 'black' or 'white' or, in binary-digital notation, '1' or '0'. In the HR mode the document is thus converted into a string of approximately 4×10^6 bits or 4 Mbits. If the SR mode is used then approximately 2 Mbits are obtained per document.

If the document being processed consists solely of text, the amount of digital information may be considerably smaller and may be input directly into the system from a word processor. A maximum of 50

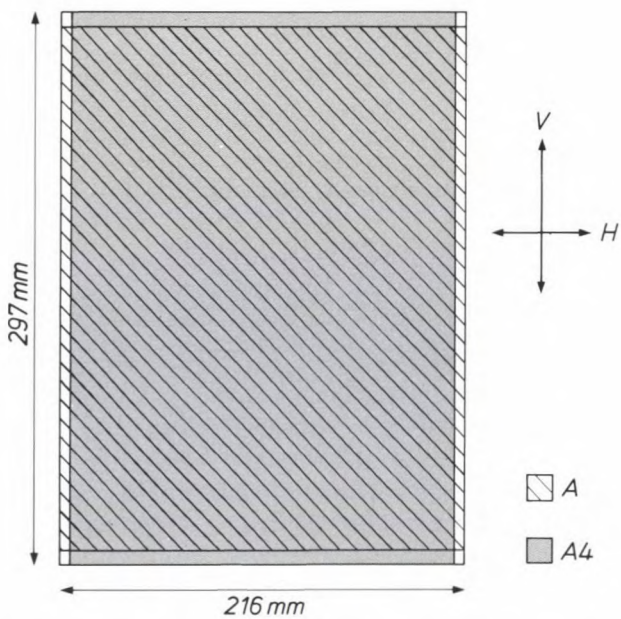


Fig. 1. The maximum format of the documents. Documents of both the European A4 format (210 × 297 mm) and the American A format (216 × 279 mm) can be processed. *V* resolution in vertical direction; 7.7 lines/mm in the HR mode, 3.85 lines/mm in the SR mode. *H* resolution in the horizontal direction, 8 pixels/mm. In the HR mode the document is divided into 2287 lines of 1728 pixels, a total of approximately 4×10^6 pixels

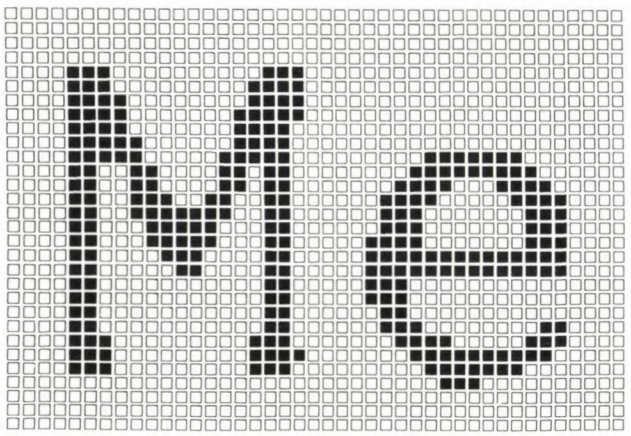


Fig. 2. Detail (magnification about 15×) of a document subdivided into approximately 4×10^6 pixels in the HR mode. The 'black' pixels are shown as filled squares, and the 'white' pixels as open squares. This shows that the details of a typed alphanumeric character are clearly visible with this fine subdivision.

lines of 80 characters must then be processed. Each character is translated into a word of 7 bits in one of the most commonly used codes, the ASCII code (American Standard Code for Information Interchange). Including one bit that can be used for checking the parity (the number of times that '1' occurs), 8 bits are required for an alphanumeric character. This gives a maximum of 4000 bytes or a total of 32 kbits for each document.

The total of 4 Mbits for the HR mode or 2 Mbits for the SR mode can be reduced by 'compressing' the number of bits. This is done by replacing strings of bits of the same value (1 or 0) by a binary word in an agreed code. We use the modified Huffman code [2], which has been designed so that the strings that occur the most frequently on average are translated into the shortest code words. The compression factor — the reduction factor for the number of bits — depends on the nature of the document that is being converted into binary form. This factor has a maximum value of 15 for the standardized CCITT document 2 (an electronic diagram drawn freehand) and a minimum of 5 for CCITT document 7 (a document consisting of oriental characters). We shall assume that the mean compression factor is 10.

To summarize, *Table 1* shows the number of bits obtained in the various ways of converting the document into digital form.

Table 1. The number of bits per document for the various ways in which the digital information can be obtained.

	Number of bits × 10^3		
	HR	SR	Alphanumeric
Alphanumeric (maximum)	—	—	32
Pictorial (compressed)	400	200	—
Pictorial (non-compressed)	4000	2000	—

The mass memory

In the Megadoc system great quantities of digital information must be stored compactly in a memory system. We decided to make use of the professional DOR disc (DOR: digital optical recording). The disc has a diameter of 30 cm and in many ways is the professional version of the well-known VLP disc [3].

The DOR disc (see *fig. 3*) that we are using at this moment consists of two toughened-glass plates separated by an outer spacer and an inner spacer with a centring hole. On the inner side of each glass plate there is a thin film of tellurium in which the digital information can be 'written'. In the VLP disc the information is pressed permanently into the disc in the factory in the form of pits of variable length and spacing, producing a frequency-modulated signal that is processed into the video signal. In the DOR disc, however, the user produces his own pattern (see *fig. 4*) with his recording device: the pattern consists of the presence or absence of holes in one of the tellurium

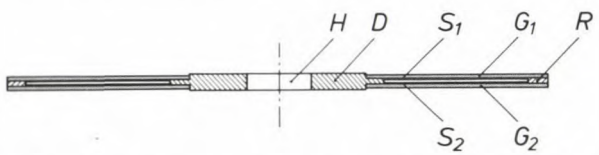


Fig. 3. Cross-section of the DOR disc. The glass plates G_1 and G_2 are kept apart by an inner spacer D with a centring hole H, and an outer spacer R. The inner surfaces of the glass plates are coated with thin films of tellurium S_1 and S_2; the digital information is written in the tellurium films. The internal space is filled with air of specified humidity.

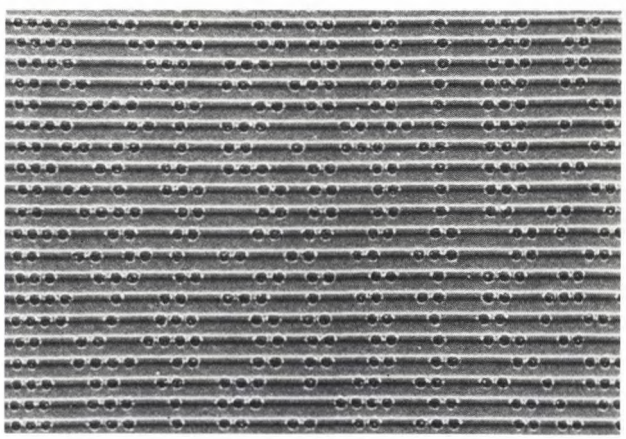


Fig. 4. Photograph (magnified about 1700×) of the information pattern in the DOR disc. The holes are 'melted' into the evaporated tellurium film by a laser beam focused on the film; the diameter of the light spot is about 1 μm. The presence or absence of a hole forms the binary-digital information relating to the document recorded in the memory. The photograph shows 20 groove turns of a part of a sector (see also fig. 5). 32 complete 360° turns correspond to the information in one single A4 document recorded in the non-compressed HR mode.

[2] U. Rothgordt, G. Aaron and G. Renelt, One-dimensional codingof black and white facsimile pictures, Acta Electronica **21**, 21-37, 1978.

[3] This is described in four articles in Philips tech. Rev. **33**, 177-193, 1973.

films. The diameter of the holes in mm and their spacing in degrees of arc (or the spacing of the places where holes could be) are constant. The holes are made by melting the film away with a laser light spot: a laser beam is focused on the film and its intensity is varied with location to produce the required pattern. The space between the two glass plates is hermetically sealed and filled with specially treated air. When the memory disc is being written or read (this can be done on either side of the disc) it rotates at a speed of 4 rev/s. This will be increased to 16 rev/s in a later version.

When the blank disc is manufactured a spiral groove with a pitch of 1.66 μm is made in each of the tellurium films. The current type of disc has 40 000 groove turns. Each side of the disc is also divided into 128 sectors (see *fig. 5*). It is possible to establish the point where the bit string of a recorded document starts, since each segment (the part of the groove turn per sector) has its own address. This address is given in coded form as part of the heading of each segment when the disc is being manufactured and indicates the location in 'polar coordinates', for example sector 27, turn 3201. A number of synchronization bits are also included in the prefabricated heading of a segment to synchronize the read/write electronics with the rotation of the disc. The number of bits that can be written into the remaining part (90%) is 1024 for a segment and 10^{10} for the entire disc (both sides).

It will be clear that the information, once written in, can no longer be changed and the recorded disc cannot be 'erased'. The DOR disc is therefore eminently suitable for storing permanent records such as files or archival material. The disc is a read-only memory (ROM) with random access. This means that any storage location can be found 'immediately' (on average within 250 msec, and this time will be reduced to 135 msec in later types of recorder unit) without the need to search a large number of successive locations as with magnetic tape. To provide a check on the effect of any possible film defects in the disc, the DRAW (DRAW: Direct Read After Write) principle is used when the bits are being written into the segments. In this process the 'write' light spot is followed at a distance of several bits by a 'read' light spot that reads back the written information immediately. A number of error-correction bits are added to the information bits being written to the segment. When a segment has been written the error-correction bits are used to correct the information read-out from this segment where possible (while the heading for the next segment is being read). If comparison with the original information still indicates errors, the information is written again in the next segment. This procedure, and the isolation of the information-containing film from the environment, leads us to predict an error probability of less than 1 in 10^{10}, possibly falling to 1 in 10^{12}, after the first ten years. Accelerated life tests on recorded DOR discs suggest that such an error probability could be a practical possibility.

The DOR disc can contain 500 000 A4 documents in alphanumeric form or 25 000 such units pictorially in the HR mode with compression. Some idea of the immense storage capacity of the DOR disc is obtained if we compare the space required by a Megadoc storage module for 64 discs with the space required by a conventional filing system. In the HR mode with compresson the Megadoc storage module may contain 1.5×10^6 A4 documents and requires the same floor space as a standard office desk. To store the equivalent amount of information on paper, however, a conventional filing system would require a cabinet about 60 metres long, 2.8 metres high and 35 cm deep.

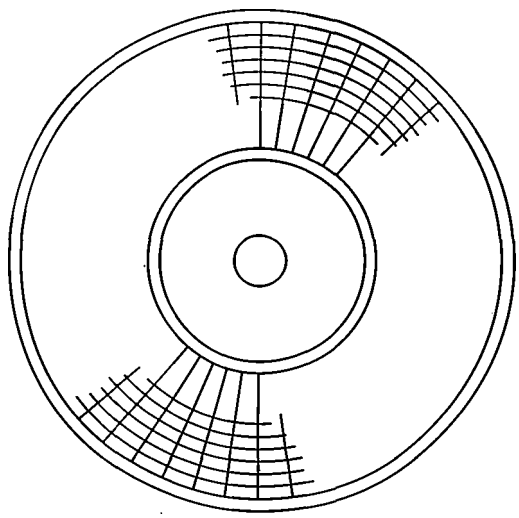


Fig. 5. Topography of the optical disc. The disc, in which each tellurium film has a spiral groove of 40 000 turns with a pitch of 1.66 μm, is subdivided into 128 sectors on each side. In the figure the pitch of the groove turns has been greatly enlarged; the angle of the sectors has been made about three times too large. The start of the bit string of a recorded document can be established in 'polar coordinates' by numbering the sectors and groove turns. In the heading of each segment (the part of the groove turn for each sector) the address is written in coded form during manufacture.

It is also interesting to compare the storage capacity of the DOR disc with that of conventional magnetic stores. The capacity of a single optical disc is equal to the capacity of 24 magnetic nine-track tapes 730 m long (2400 feet) each with 640 bits/cm (1600 bits/inch), i.e. about 17 km of magnetic tape. The information density (or packing density) of a 'floppy disc' is about a hundredth of that of the optical memory disc.

The system and its modules

Fig. 6 is a block diagram of the complete Megadoc system illustrating the various modules that form part of the system. The entire system is controlled by a minicomputer; we use a Philips P857. Two *MHD* (moving head disc) peripheral units, made by another manufacturer, are connected to the computer for information storage on magnetic discs, each with a software of most interest to the user. The DOC-MAIL program segment controls the input of documents into the system and their distribution to the users, DOC-TRACK creates, stores and searches for document descriptions while DOC-HAND controls the transport of documents from peripheral to peripheral. Access to these three program segments is controlled by the segment DOC-MAIN.

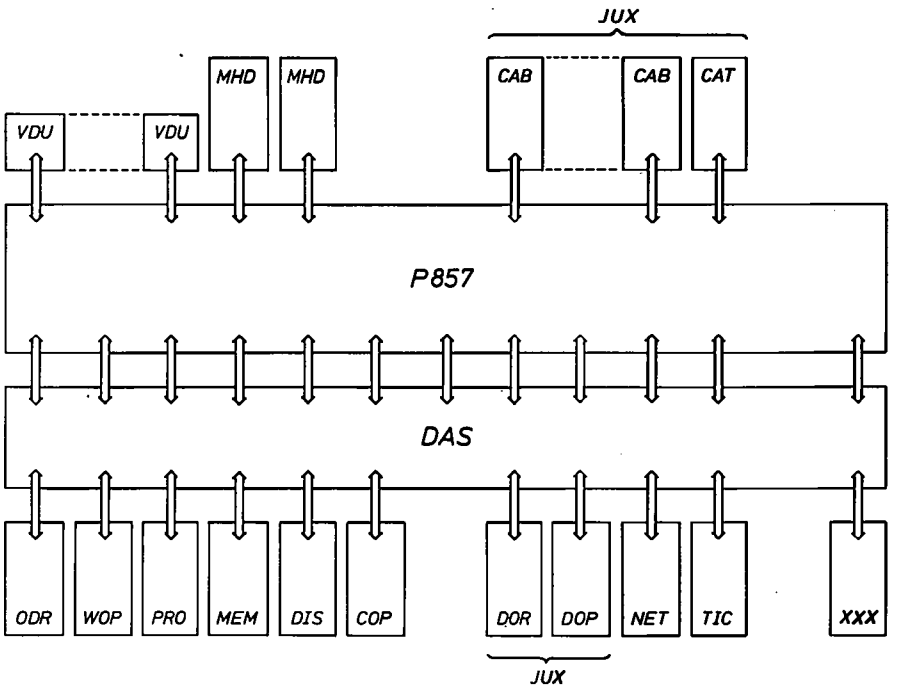


Fig. 6. Block diagram of the Megadoc system, showing the various modules of the system. The system is controlled by the Philips P857 minicomputer. Two *MHD* peripheral units are included in which data can be stored externally on magnetic discs. *VDU* video display units (maximum 16) with screen and keyboard enabling the user to communicate with the system. *CAB* modules that can each store and retrieve 64 optical discs. *CAT* disc-transport system. *JUX* combination of *CAB*, *CAT*, *DOR* and/or *DOP*. *DAS* switching network, which controls the information flow between the various peripheral units. *ODR* document reader. *WOP* word processor. *PRO* unit for 'compressing' and 'decompressing' the binary-digital information. *MEM* semiconductor memory in which the information from a single document can be stored temporarily. *DIS* combination of document display and refresh memory. *COP* unit for producing hard copies. *DOR* reader/writer for optical discs. *DOP* reader for optical discs. *NET* provision for connection to the public telephone system or future information networks. *TIC* unit for converting digital information from alphanumeric into pictorial form. *XXX* provision for extension.

total capacity of 40×10^6 bytes. These form the external storage capacity of the computer. The internal (semiconductor) memory capacity is 128×10^3 words of 16 bits. In addition, a maximum of 16 video display units (*VDU*) can be added, each with display and keyboard, which can communicate with the system independently of one another. A maximum of 16 peripheral units for handling documents can be connected to the system. One *VDU*, plus one or more peripherals, can form a 'work station'.

Fig. 7 is a diagram showing the parts of the system

The flow of information between the various document-handling peripherals goes through the electronic data switch *DAS*. This data switch was developed specially for Megadoc at the Geldrop Project Centre. Its development was necessary because a maximum handling time of one second per document has been specified for the system. The binary information must therefore be transferred in less than a second at the very high effective rate of 4 Mbits/s. Such high transfer rates cannot be processed by the P857 minicomputer for a relatively long period of time, so that

the flow of information has to by-pass the computer. *Fig. 8* is a schematic representation of the information traffic in the data switch. The entire system is controlled by a Signetics 2650 microprocessor, which is controlled in turn by the central minicomputer. The microprocessor controls the traffic between a maximum of six pairs of peripherals that can communicate with one another simultaneously through the channels C_0 to C_5. The bit stream is subdivided into bytes that are transferred simultaneously. Each channel consists of ten connections, eight for transmitting parallel bytes and two for 'handshaking' signals; these handshaking signals are sent between transmitting and receiving units to indicate when each is ready. The peripherals are each connected to switching units (S_0 to S_F) of *DAS* by fifteen connections, eight for the information bytes and seven for check signals, two of these for the 'handshaking'. The information bytes are transmitted with the aid of the check signals under the control of the switching units S_0 to S_F, making use of a standardized protocol. This protocol is part of a 'general-purpose interface', for transferring parallel binary-digital information, in which the electronic control units I_0 to I_F form part of the P857 minicomputer.

The standard version of *DAS* should not be further than 25 m away from a peripheral unit. If a specially extended version of the data switch, *DAS-EXD* (*DAS* extended) is used, this distance can be increased to 400 m. This is achieved by changing the procedure for the 'handshaking' and by transferring the bits sequentially by baseband data transmission (one pulse per bit) along a single coaxial cable at a rate of 20 Mbits/s.

The following peripherals can be connected to the system (see fig. 6).

ODR (original document reader). This unit can scan a document optically in one second and process it into one of the numbers of bits shown in Table I for the HR or SR mode. It was developed entirely by the Geldrop Project Centre and will be discussed in greater detail below.

MEM (transfer memory). This is a conventional buffer memory, based on IC semiconductor circuits, in which the information can be temporarily stored by a microprocessor control system that we have developed. The memory units are standardized and have a total capacity of 8 Mbits. Once again the maximum input and output rates are 4 Mbits/s.

DIS (document display). This is a combination of a high-resolution display and a refresh memory. Any document present in the Megadoc system can be called up to appear on the document display. At a frequency of 47 Hz the refresh memory provides the

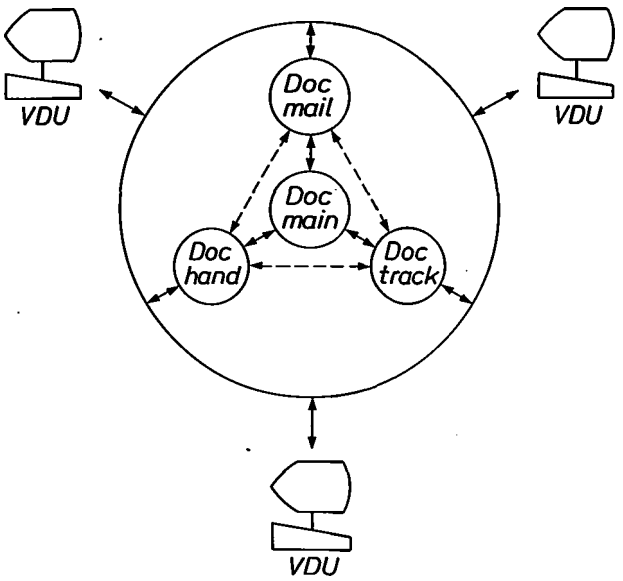


Fig. 7. A diagram showing the parts of the software of most interest to the user. Access to the three program segments DOC-MAIL (input and distribution of documents), DOC-TRACK (creation, storage and searching for document descriptions) and DOC-HAND (transport of documents from peripheral to peripheral) is monitored by the program segment DOC-MAIN. The terminals *VDU* can be used to call up the various program segments and to communicate with the system.

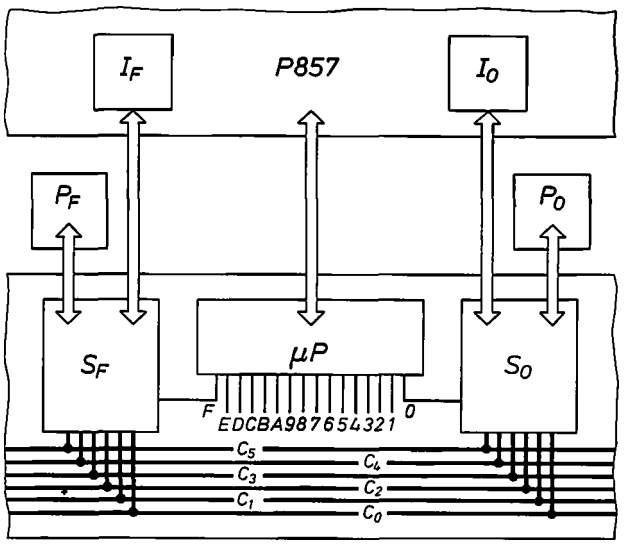


Fig. 8. The organization of the data switch *DAS*. A maximum of 16 peripheral units P_0 to P_F (0 to F, 16 digit or hexadecimal indication for 0 to 15) can be linked together by means of channels C_0 to C_5. A maximum of 6 pairs of peripheral units can therefore be connected together simultaneously. Each peripheral is connected to its own switching unit S_0, S_1 ... or S_F (they are all identical), thus effecting the channel link selected by the microprocessor μP (Signetics 2650). The microprocessor has 16 outputs *0* to *F* and is under the control of the central minicomputer P857. The binary-digital information is transferred at an effective rate of approximately 4 Mbits/s. The bits are transferred in parallel bytes from peripheral unit to switching unit under the control of the switching units, in a standard protocol, accompanied by seven check signals, including two for 'handshaking'. This protocol is part of a general-purpose interface for transferring parallel binary-digital information, in which the electronic control units I_0 to I_F form part of the P857 minicomputer. Channels C_0 to C_5 each consist of ten connections, eight for passing on the information bits and two for the check signals for the 'handshaking'.

full information from a single image in the HR mode for the video signal entering the display electronics. This is also one of our own developments and will be discussed in greater detail later.

DOR (digital optical recorder). This unit can be used to write the digital information to the disc and to read it from the disc. This equipment is now being developed by Philips.

DOP (digital optical player). This unit is only able to read the optical disc. This equipment is another development from the Project Centre.

JUX (jukebox modular system). This is a combination of storage modules *CAB* (cassette box) for 64 memory discs, disc-transport systems *CAT* (cassette transport) and one or more *DOP* readers or *DOR* read/write units. All these units — except *DOR* — are our own developments; we shall discuss the *CAB* disc-storage module in greater detail later.

PRO (picture processor). This equipment, which is still under development at the Project Centre, will be used to provide the desired compression of the number of bits per document, as mentioned earlier, and is also able to reconstitute the original bit string.

NET (network interface). This unit consists of standardized modems with a microprocessor control system of our own development; it provides the connection to the public telephone system or to public information networks to be set up in the future (for example the DN1 network being developed jointly by the European national telephone services). This enables different Megadoc systems to communicate with one another.

COP (hard-copy unit). This enables a hard copy to be obtained of a document stored in the system. It includes a commercially available copier that requires 15 seconds per A4 side, however, so it does not achieve the processing time of one second specified for Megadoc.

WOP (word processor). This is a Philips P5002 word processor. This machine can be used to enter a document into the system in alphanumeric form.

TIC (text image convertor). This equipment is being developed at the Project Centre. It converts information recorded in alphanumeric form by *WOP* into a pictorial form suitable for *DIS*.

Some modules in detail

The original document reader (ODR)

Fig. 9 shows a photograph of the complete document reader. The document is placed under the plastic flap on top of the unit, with the text downwards; it is read in one second. The reader consists of optical, mechanical and electronic assemblies; the optical and mechanical assemblies are integrated and mounted in the upper part (see *fig. 10*); the electronic assembly is mounted in the base.

Fig. 9. The document reader *ODR*. The document is placed under the plastic flap with the text facing downwards and is then scanned optically in less than one second.

Fig. 11 is a diagram showing the path of the beam. The two light sources L_1 and L_2, the screens with slotted aperture A and the mirror M_1 move at a scanning velocity V in a direction parallel to the document D resting on a glass plate. The mirrors M_2 and M_3 move in the same direction at a velocity of $V/2$. A sharp image of a narrow strip of the document ('the line') is obtained on the linear detector S. The 2:1 ratio of the velocities of M_1 and the combination of M_2 and M_3 ensures that the image of the line of the document at S remains sharp. L_1 and L_2 are fluorescent lamps specially developed by Philips. The fluorescent layer in these lamps has been partially removed and the lamps also contain an internal reflector to give very bright illumination of the document. The detector S consists of a row of 1728 light-sensitive diodes with a centre-to-centre distance of 16 μm and a light-sensitive area of 16×16 μm. The magnification of the lens *Le* is such that the dimensions of the row of diodes correspond to the area of the row of pixels (the 'line') being scanned on the document. The mixture of the phosphors in the lamps L_1 and L_2 has been selected to give the most suitable combination of spectral distribution of the illumination and spectral sensitivity of the detector for the most natural impression when the document is reproduced later.

Fig. 12 illustrates the design of the mechanical assembly. The mirror M_1 and the light sources L_1 and L_2 are mounted on the carriage *ILS*, which is driven by the drum *Dr* and a steel cable C_1. The mirrors M_2 and M_3 are mounted on the mirror carriage *MIS*. The two carriages are connected to one another by a second cable C_2, which is taken round the pulley W_1 mounted on the carriage *MIS*. One end of C_2 is attached to *ILS* and the other end is attached to the frame of the equipment by an adjusting mechanism *F*. This design ensures that the velocities of the carriages *ILS* and *MIS* are in the ratio 2:1. This can be seen by comparing the cable C_2 with the 'folded' light path via the mirrors M_2 and M_3. A third cable C_3 keeps all the cables in the mechanism taut by means of the helical tension spring *T* and eliminates backlash. The adjusting mechanism *F* is included to provide fine focusing in addition to the coarse adjustment given by a focusing device built into the lens mount. The drive drum *Dr* and the carriages are brought up to speed within 0.1 s by a drive motor with a disc rotor to give a low moment of inertia. A control circuit that includes a tachogenerator connected to the motor then keeps the speed accurately constant for approximately one second; the drum and carriages are then finally

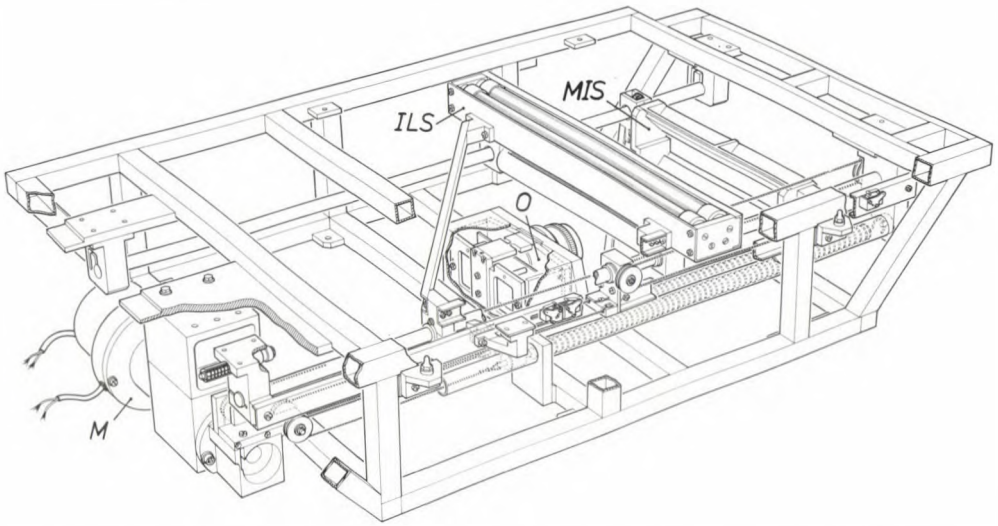


Fig. 10. The optical and mechanical assemblies of the *ODR* document reader. The most important components are the illumination carriage *ILS*, the mirror carriage *MIS*, their common drive *M* and the objective plus detector *O*. The operating principle is illustrated in more detail in figs 11 and 12.

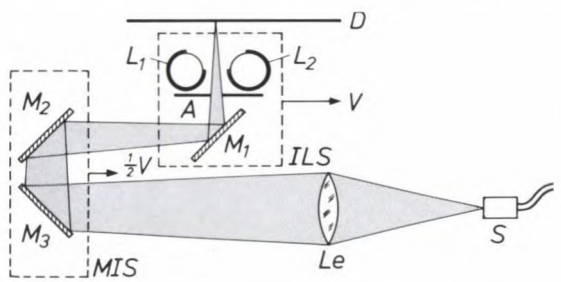


Fig. 11. Diagram illustrating the optics of the document reader. A sharp image of a line (perpendicular to the plane of the drawing) of the document *D* is produced at the detector *S* via the slot *A*, the mirrors M_1, M_2 and M_3 and the lens *Le*. The detector consists of a row of 1728 light-sensitive diodes with a centre-to-centre spacing of 16 μm and an area of 16×16 μm. L_1 and L_2 are fluorescent lamps to provide intense local illumination of the document. L_1, L_2 and M_1 are attached to the illumination carriage *ILS*; M_2 and M_3 are attached to the mirror carriage *MIS*. Since the speed of *MIS* is only half that of the carriage *ILS* a sharp image of the document line is retained on detector *S* while the document is being scanned.

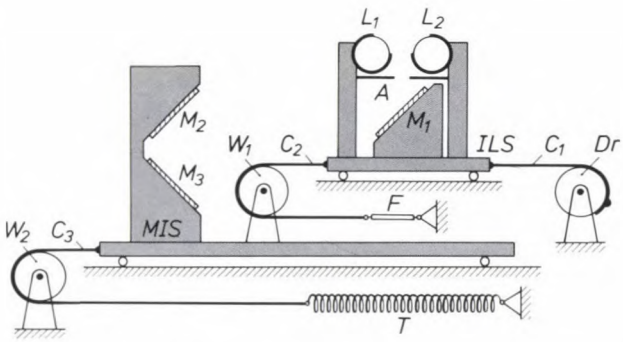


Fig. 12. Diagram illustrating the mechanics of the document reader. The carriage *ILS* is driven by the drum *Dr* and cable C_1. A second cable C_2 is connected to *ILS* at one end and to the frame of the unit at the other by an adjusting mechanism *F*. C_2 causes carriage *MIS* to move via pulley W_1. All the cables in the mechanism are kept taut and backlash is eliminated by a helical tension spring *T*, attached to a third cable C_3 that goes round the pulley W_2 attached to the frame. The adjusting mechanism *F* can be used to improve the focus. Cable C_2 can be compared with the 'folded' light path via mirrors M_2 and M_3, see fig. 11. The velocities of the carriages *ILS* and *MIS* are therefore in the ratio of 2:1. The other symbols have the same significance as in fig. 11.

slowed down again within 0.1 s. The accurate control of the speeds of the carriages is necessary to obtain a scan with the correct number of lines and a constant resolution over the entire length, since the electronic circuits that control the motor are not synchronized with the sampling of the detector.

The detector *S* consists of an integrated circuit that contains not only the 1728 photodiodes already mentioned but also two shift-registers, each consisting of a group of 864 charge-coupled devices (CCDs), and 2×864 electronic switches. During the adjustable integration time — whose maximum value is equal to the scanning time of 1/2287 s for one line — the photoelectric current produces a charge in the capacitor formed by each diode. A voltage pulse provided by the control electronics ensures that the charges from all the 'even' diodes go via the electronic switches to one of the shift registers and that the charges from all the 'odd' diodes go to the other shift register. The same control electronics ensures that all the elements of the two shift registers are then 'read' in sequence and converted into an analog video signal in which the height of each successive voltage pulse corresponds to the grey level of the appropriate pixel. A level discriminator then processes the analog video signal to give a binary video signal in which a high voltage level corresponds to a 'black' pixel and a low voltage level corresponds to a 'white' one. Finally, this digital video signal is processed by a serial-to-parallel converter into bytes that are then transferred further by *DAS* in accordance with the standardized protocol.

The document display with refresh memory (DIS)

Fig. 13 shows the document display with refresh memory. The display unit is mounted on the desk, the angle can be adjusted to suit the user; the electronic circuits for the refresh memory are located in the cabinet underneath.

The cathode-ray tube has been selected as the means of displaying the document in the Megadoc system since at the moment it is the only device that can provide sufficient resolution and a large enough picture. The picture tube was specially designed by Philips for applications where high resolution is required. The tube is mounted with its long side vertical, it has a picture diagonal of 38 cm (15″) and an aspect ratio of 4:3, so that the document can be displayed on a scale of 1:1. The angle of deflection is 70° and the neck diameter is 28 mm. To obtain a positive and natural image of the document a white luminescent phosphor of the P4 type is used. In the HR mode 2287 horizontal lines are written per picture, and fields of odd and even lines always appear in succes-

Fig. 13. The document-display screen with refresh memory. The angle of the document-display screen on the desk can be varied; the electronic circuits for the refresh memory are located below the desk.

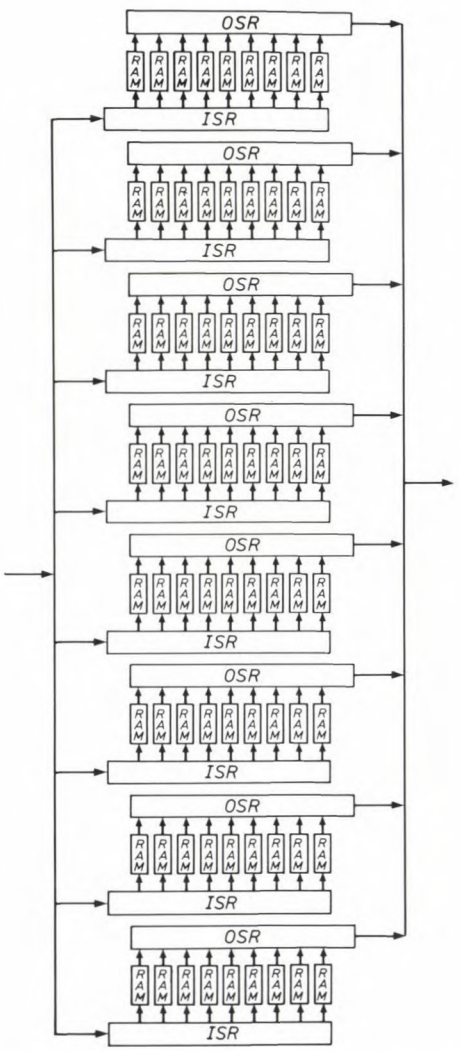


Fig. 14. Block diagram of one of the four memory blocks *MEB* of the refresh memory. *ISR* input shift registers, *OSR* output shift registers. *RAM* random-access memory circuits that can each contain $128 \times 128 \approx 16$ kbits.

sion (interlaced 2:1), so that the actual frame frequency of 47 Hz appears to be at twice this frequency and flicker is avoided. In the SR mode the interlaced fields are made up of the same 1143 lines. The combination of deflection coils has been specially designed by Philips for use with this tube.

The function of the refresh memory is to present the information for a complete frame in the HR mode at a frequency of 47 Hz to the monitor electronics 9 RAM memory circuits and each RAM circuit may contain $128 \times 128 = 16384$ bits. The total capacity of the refresh memory is therefore $4 \times 8 \times 9 \times 128 \times 128 = 4718592$ bits. Since, as we saw earlier, $2287 \times 1728 = 3951936$ bits are required for a pictorial display of a document in the HR mode (non-compressed) the storage capacity of the memory is amply sufficient.

Fig. 14 gives the block diagram of one of the four *MEB* memory blocks. Each row of 9 RAM circuits

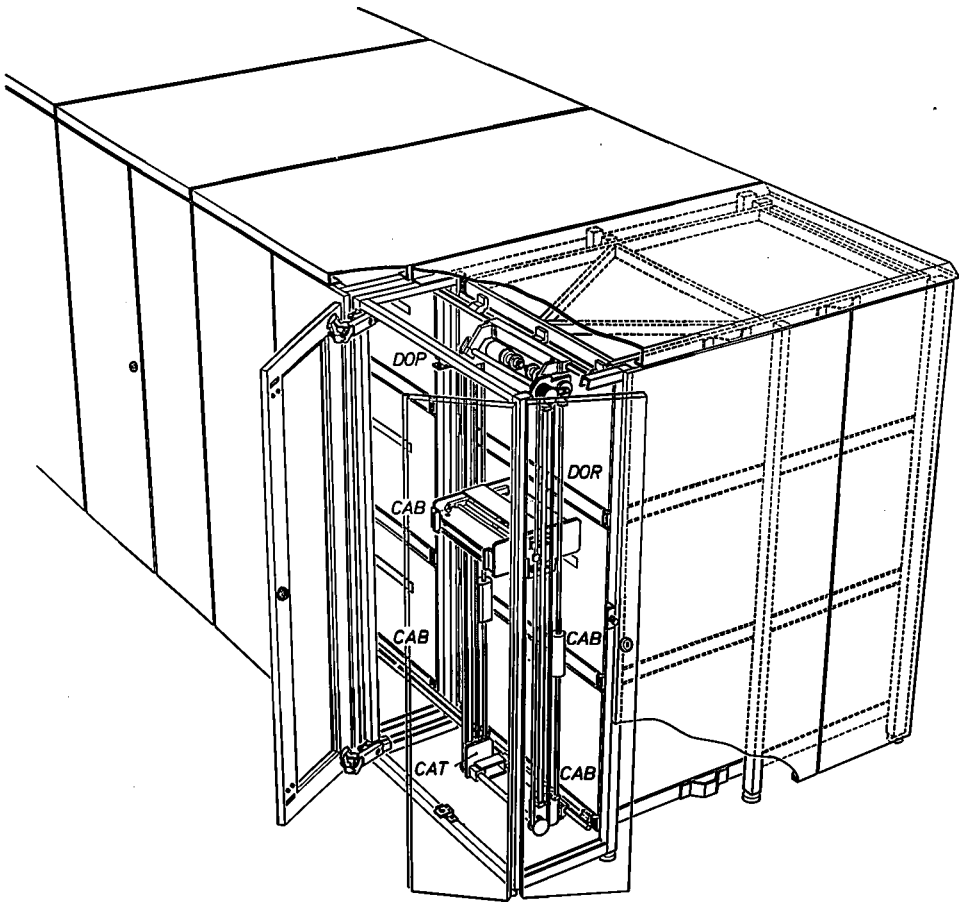


Fig. 15. Diagram of a possible configuration of the *JUX* storage system for optical discs, viewed from the rear. The *CAB* storage modules are linked to one another and to a *DOR* reader/writer and a *DOP* reader by means of the cassette-transport system *CAT* consisting of a vertical and a horizontal rail system.

located inside the housing of the picture tube. This means that 4 Mbits must be continuously presented at this frequency. Because of the time required for the flyback of the electron beam in the cathode-ray tube the refresh memory must therefore be capable of operating at a bit rate of about 255 Mbits/s. Since the RAM semiconductor memory circuits have a minimum cycle time of 375 ns, several RAM memory circuits have to be used in parallel. As well as a number of control circuits the refresh memory therefore includes four memory blocks (*MEB1, 2, 3* and *4*). Each memory block consists of eight rows each with has its own input shift register (*ISR*) and output shift register (*OSR*). The bit strings of the odd lines are written to memory blocks *MEB1* and *2* and those of the even lines are written to blocks *MEB3* and *4*. This is done by feeding the bits for each byte in parallel to the input shift registers. In this way the first 16 bits of the first two bytes of the first line are simultaneously entered into the 16 input shift registers of *MEB1* and *2*, into the first place in each register. When the second string of 16 bits of the first line arrives in the input shift registers, these in their turn enter the first place and the bits that are already there are shifted into the

second place. This is repeated until all nine places of the input shift registers have been filled. When the tenth string of 16 bits arrives the $9 \times 16 = 144$ bits already present are simultaneously transferred into the RAM circuits, to the same place in each circuit, so that all these bits are allocated the same address. This procedure is repeated twelve times until all the $12 \times 144 = 1728$ bits of the first line have been stored in memory blocks *MEB1* and *2*. The bits of the second line are then entered in *MEB3* and *4* in a similar manner and so on until the information from the entire frame has been processed. The addresses relating to the locations of the bits in the RAM circuits are generated and monitored by one of the control circuits mentioned above.

With the quite acceptable cycle time of 560 ns, corresponding to a frequency of $\frac{1}{144} \times 255$ Mbits/s, the binary information is retrieved from the RAM circuits line by line so as to yield a complete frame at a frequency of 47 Hz. The 16 bit parallel bytes are strung together again to form a sequential binary video signal. After further treatment in the monitor electronics this video signal is applied to the cathode of the cathode-ray tube.

The monitor electronics also generates the sawtooth currents for the horizontal and vertical deflection. These currents are supplied to the horizontal- and vertical-deflection coils. The sawtooth currents are generated via the control circuits of the refresh memory, with synchronization from the clock-pulse generator. The monitor electronics also contains a circuit for the 'dynamic focusing'. This circuit ensures that once the beam has been focused in the centre of the screen it also remains in focus over the rest of the display screen.

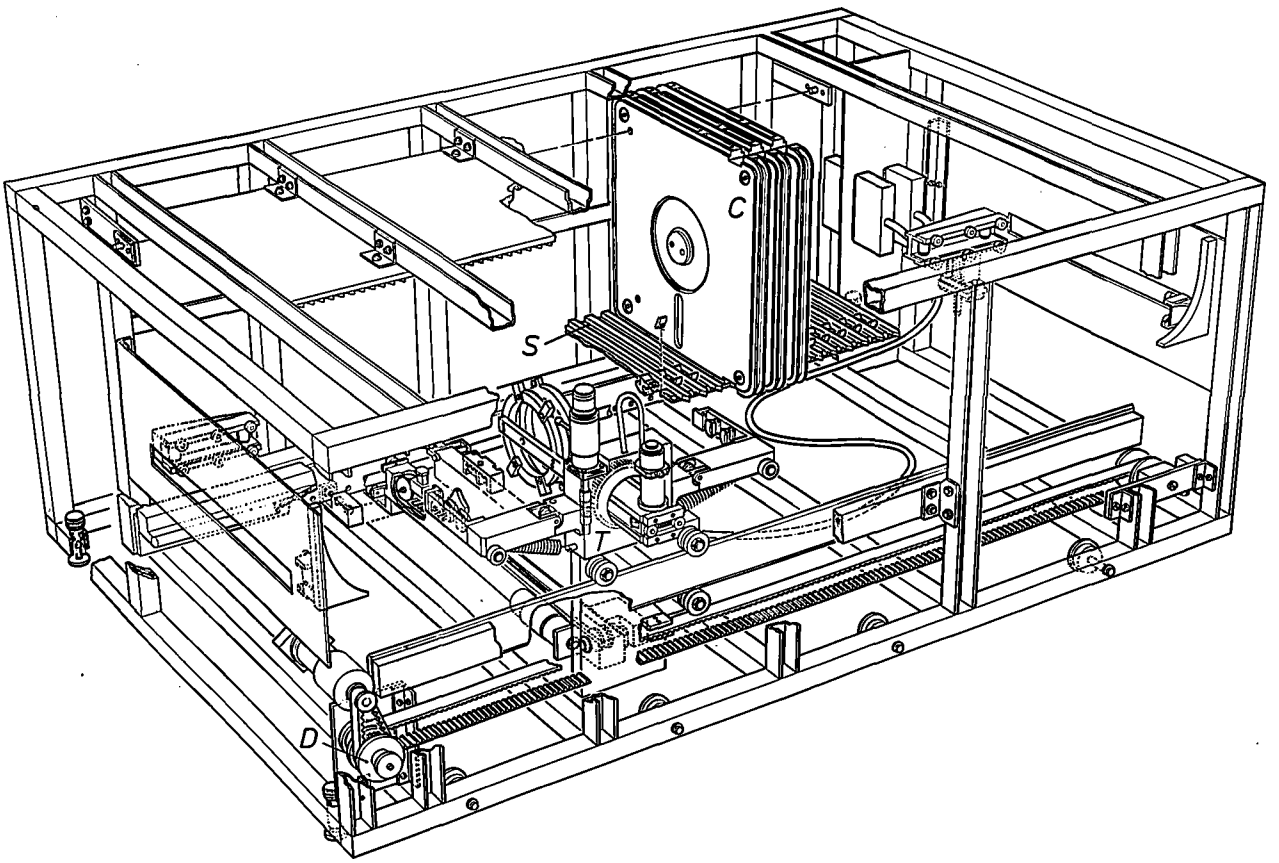


Fig. 16. The inside of a *CAB* storage module for 64 discs. *C* cassettes with optical discs. *S* lower guide rails for the cassettes. *T* cassette-transport mechanism. *D* drive for the transport mechanism, which employs toothed belts.

Storage module for optical discs (CAB)

The optical discs, each housed in a cassette, are stored in a system largely built up from *CAB* storage modules. As mentioned earlier, several *CAB*s, together with a *CAT* cassette-transport system and one or more *DOP* readers or *DOR* read/write units, form a complete *JUX* system. An example is shown in *fig. 15*.

We shall now take a closer look at the construction and operating principle of the *CAB* storage module; see *fig. 16*. Up to 64 disc cassettes (see *fig. 17*) are placed next to one another in grooved guide rails. The plastic cassettes are in two halves, which are held slightly apart while the disc is being written or read, so that the rotating disc does not come into contact with the cassette. The cassettes have a slot *Sl* with a sliding seal to give access to the write or read head, and two recesses *R* enabling the cassette to be extracted from its guide rails. Alignment holes *H* through the corners of the cassettes are provided so that the alignment of the cassettes can be checked by means of a lamp and a photocell.

Fig. 18 illustrates the cassette-transport mechanism for movement in a direction parallel to the row of cassettes. The mechanism can remove a cassette plus disc from its guide rails and transfer it to the point where the cassette-transport system *CAT* takes over. The mechanism contains a toothed belt *B* with a stud *St* attached to it; the stud passes freely through the recesses *R* as the mechanism moves along the row of cassettes. When the desired location has been reached the toothed belt is actuated and the cassette slides out of its guide rails. The cassette clamp *C*, which is now in a vertical position, then closes, raises the cassette slightly and turns it through 90° so that the side of the disc to be read or written is at the bottom. At about the same time the transport mechanism is approaching the end of its run; the clamp then opens to transfer the cassette to the cassette-transport system *CAT* (see fig. 15).

Fig. 19 shows how the position of the cassette-transport mechanism is determined with respect to the cassette guide rails. To determine the position, a shaft-position indicator (not shown in the figure), based on a rotary Gray code, is mounted on the motor shaft that drives the toothed belt that moves the cassette-transport mechanism in the longitudinal direction (see fig. 16). The final determination of the position of the mechanism is achieved by means of a perforated scale *Sc*, which enables an analog positioning signal to be obtained by means of a detector *D*, consisting of two halves, and an LED (light-emitting diode) *L*. The desired position has been reached when the two halves of the detector give the same photoelectric current. The figure also shows how the transport mechanism can determine whether or not a cas-

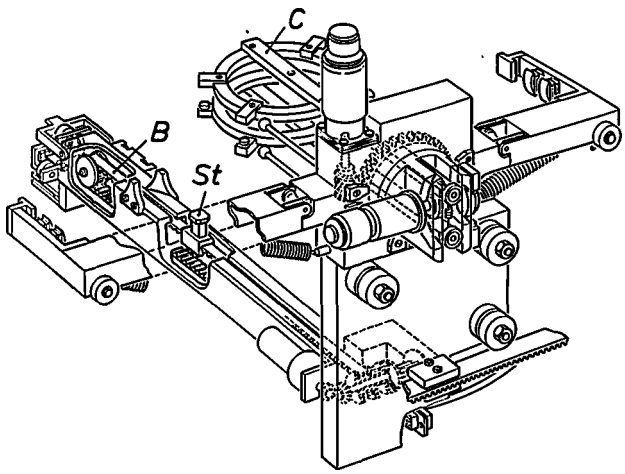


Fig. 18. Cassette-transport mechanism. The mechanism can be moved parallel to the row of disc cassettes in the storage module; the stud *St* moves freely through the recesses *R* (see fig. 17) in these cassettes. When the desired cassette is reached, the toothed belt *B* is activated and the stud *St* pulls the cassette out from its guide rails. The clamp *C* which is now in a vertical position, then tightens. At about the same time as the clamp is rotating back into the horizontal position the transport mechanism is moving towards the end of its run and the cassette is delivered to the transport system *CAT* (see fig. 15) by opening the clamp.

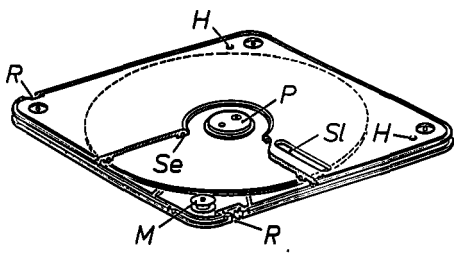


Fig. 17. The optical disc cassette. *Sl* slot for optical reading from or writing to the disc. *P* centring stud enabling the disc to be centred on the spindle of *DOR* or *DOP* during reading or writing. *R* recesses for extracting the cassette from its guide rails by the transport mechanism. *H* holes for periodic alignment checks to verify that the cassettes are still in their correct positions in the guide rails (see fig. 16). *M* membrane springs. The two halves of the cassette are connected by these springs. When the two halves have been slightly separated the disc can be read from or written to without the cassette coming into contact with the rotating disc. *Se* sealing ring, to give a dust-tight seal for the disc in the cassette. A sliding seal (not shown) is also provided for slot *Sl*.

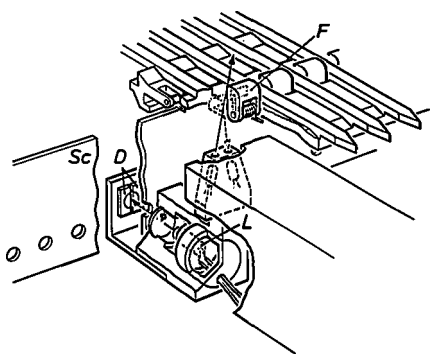


Fig. 19. Positioning of the cassette-transport mechanism. *Sc* scale with perforations for accurately determining the position, with the aid of a light-emitting diode (LED) *L* and a light-sensitive detector *D* in two halves. The spacing of the holes in the scale is equal to the spacing of the guide rails. The hinged flap *F* indicates whether there is a cassette in a guide rail. The flap has a special reflecting surface and an LED-photodiode combination is mounted on the cassette-transport mechanism.

sette is present in its guide rails. The slot of each lower guide rail has a hinged flap *F*, which is held in a vertical position by a spring when the cassette is not there. The insertion of a cassette causes the flap to take up a horizontal position. An LED-photodiode combination attached to the transport mechanism detects the position of the flap, which has a special reflecting surface for this purpose.

The cassette-transport mechanism and its drive also contain various shaft-position indicators, tachogenerators, microswitches and detectors that monitor the positions of the mechanism and check for the presence of the cassette. The transport instructions given by the central computer can thus be executed without error. A Signetics 2650 microprocessor therefore provides the local control and is incorporated in a control module that forms part of the *JUX* system.

The software

The information obtained from the document and stored in the Megadoc system can be divided into two distinct parts. The first part is the binary-digital and mainly pictorial information from the document itself. As explained earlier, this is stored at a given address in one of the optical discs. The other part is the description of the document and contains data such as the name of the author, the title, miscellaneous data and certain keywords. This information is stored in a data base on the magnetic discs of one of the *MHD* peripheral units. This descriptive information can be used for finding documents when the system is supplied with characteristic data such as the date, the author's name or certain keywords. The documents description is recorded in the data base under a document number issued by the system. The document number is directly linked in the data base to the address of the relevant information in one of the optical discs. A quarter of the external storage capacity of 40×10^6 bytes of the first of the two peripheral units *MHD* is reserved for the data base and is equivalent to 33 000 document descriptions. In the remainder, a capacity of 20×10^6 bytes is intended as a reserve for the storage of document descriptions. The total of almost 100 000 document descriptions is just about enough if a single *CAB* storage unit is added to the system, assuming that there is one description per average document of 15 pages. The storage capacity of the document descriptions may be increased in the future by using optical discs instead of magnetic discs.

The standard software supplied with the P857 minicomputer permits different applications, each consisting of one or more programs, to be active simultaneously. This is done under the 'supervision' of the 'monitoring program' MAS (multi-application system), which has a size of 30×10^3 words of 16 bits and is stored externally in the second peripheral unit *MHD*. This program controls the progress of the applications by allocating computer time and areas of the memory. MAS also ensures that the programs called by the applications are numbered in sequence depending on the degree of urgency. MAS does this by continuously updating the list of programs called by the applications. As soon as a program has completed its allotted tasks it is removed from the list.

The SYS (software system) application, which has been specially developed for Megadoc, is written in assembly language — a computer language whose level is only slightly higher than the 'machine language'. SYS has a coverage of almost 30×10^3 words of 16 bits and is stored externally in the first peripheral unit *MHD*. It consists of two program packages: DSX (document switching and exchange) and DSR (document storage and retrieval). The DSX programs (there are about 30) are responsible for document transport from peripheral unit to peripheral unit and have no direct communication with the user. The DSR program package contains only one program, which controls the communication of the user with the system and consists in turn of the program segments DOC-MAIL, DOC-TRACK and DOC-HAND and the access program segment DOC-MAIN, all of which were mentioned earlier. The execution of this program is determined by the user, who is repeatedly asked to select options. He has to answer questions appearing on his *VDU* screen, either by pressing one of the keys F1, F2, F3, F4 or F5 on his keyboard or by typing in text or figures. At each stage the screen will indicate the significance of these five keys; the group of five possible selections is known as the 'menu'. *Fig. 20* shows the complete DSR software in the form of a block diagram in which each block represents one selection. First of all the user has to make a choice from the three program segments. He can then follow a given path in the diagram by answering the questions from the selected program segment.

The main activities that the three program segments can carry out are:

— for DOC-MAIL

storing the pictorial information from a document on an optical disc,

distributing documents with appropriate descriptions to the various users of the system (mailbox function);

— for DOC-TRACK

obtaining document-description data from the user,

searching for a document using the data file formed

from the existing document descriptions,

providing lists of authors or keywords;

— for DOC-HAND

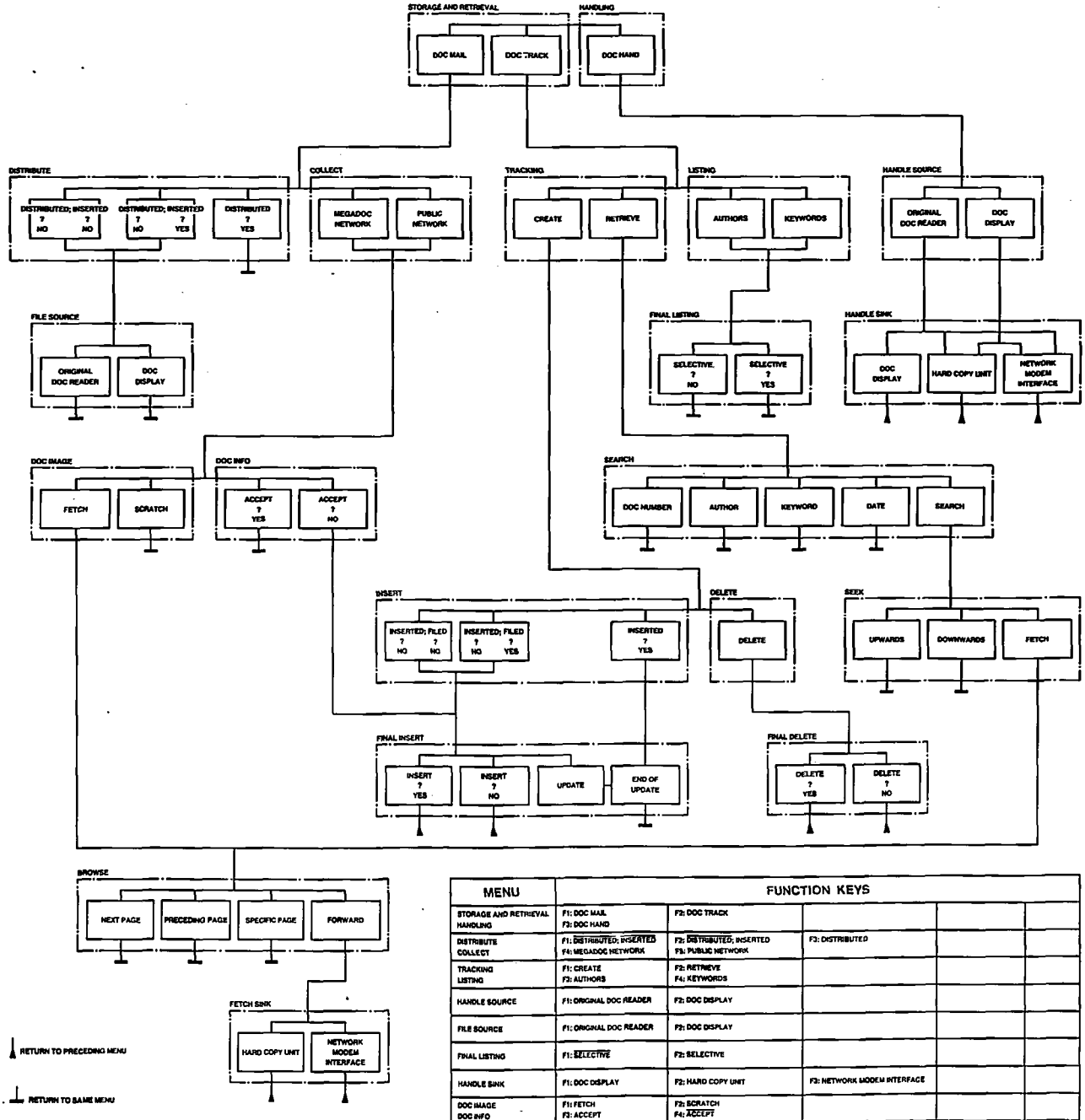
selecting a peripheral unit for transmitting a document,

selecting a peripheral unit for receiving a document;

DSX programs that execute the document transport are then activated.

We shall now look at several examples of transactions made by the user; these follow various paths in the diagram of fig. 20.

From her work station, consisting of *VDU*, *DIS* and *ODR*, the secretary distributes the incoming mail with the aid of the DOC-MAIL program segment. The system responds to the selection DISTRIBUTED? NO, INSERTED? NO by first of all producing on the



MENU	FUNCTION KEYS				
STORAGE AND RETRIEVAL HANDLING	F1: DOC MAIL F3: DOC HAND	F2: DOC TRACK			
DISTRIBUTE COLLECT	F1: DISTRIBUTED; INSERTED F4: MEGADOC NETWORK	F2: DISTRIBUTED; INSERTED F5: PUBLIC NETWORK	F3: DISTRIBUTED		
TRACKING LISTING	F1: CREATE F3: AUTHORS	F2: RETRIEVE F4: KEYWORDS			
HANDLE SOURCE	F1: ORIGINAL DOC READER	F2: DOC DISPLAY			
FILE SOURCE	F1: ORIGINAL DOC READER	F2: DOC DISPLAY			
FINAL LISTING	F1: SELECTIVE	F2: SELECTIVE			
HANDLE SINK	F1: DOC DISPLAY	F2: HARD COPY UNIT	F3: NETWORK MODEM INTERFACE		
DOC IMAGE DOC INFO	F1: FETCH F3: ACCEPT	F2: SCRATCH F4: ACCEPT			
SEARCH	F1: DOC NUMBER	F2: AUTHOR	F3: KEYWORD	F4: DATE	F5: SEARCH
INSERT DELETE	F1: INSERTED; FILED F4: DELETE	F2: INSERTED; FILED	F3: INSERTED		
SEEK	F1: UPWARDS	F2: DOWNWARDS	F3: FETCH		
FINAL INSERT	F1: INSERT	F2: INSERT	F3: UPDATE	F4: END OF UPDATE	
FINAL DELETE	F1: DELETE	F2: DELETE			
BROWSE	F1: NEXT PAGE	F2: PRECEDING PAGE	F3: SPECIFIC PAGE	F4: FORWARD	
FETCH SINK	F1: HARD COPY UNIT	F2: NETWORK MODEM INTERFACE			

Fig. 20. Block diagram of the one program of the DSR (document storage and retrieval) program package. Each block represents a selection that the user can make from the 'menu' displayed on the screen of his *VDU*. The top row of blocks offers the user the choice between the program segments DOC-MAIL, DOC-TRACK and DOC-HAND. Answering the questions put to him by the system takes the user along a certain path in the diagram.

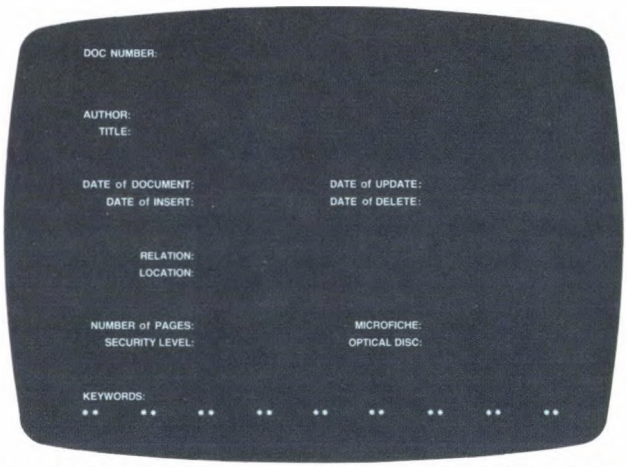


Fig. 21. The layout of the document description as shown on the *VDU* display. Most of the items have to be entered by the user, some of the others such as the document number and the number of pages are generated by the system.

screen a list of the names of the people served by the system. The secretary selects the names of the people on the list who will receive the document now being input. The system then requests the desired security level of the document. When she has given this, the secretary can input the document into the system by the *ODR* scanning unit; up to 8 seconds per page can be taken for this operation. The secretary may have already provided the document with a full document description or, as we shall assume here, she may leave this to the user.

At his own work station, consisting of *VDU* and *DIS*, the user examines the post that has come in for him. He does this also by means of the DOC-MAIL program segment, by selecting MEGADOC NETWORK. The system first of all finds out whether any documents have been entered for that particular user. If the answer is no, the message NO MORE DOCUMENTS PRESENT IN POSTBOX appears. If there is a document intended for him, the document description (for layout see *fig. 21*) appears on his *VDU* screen and the first page of the document appears on the document display *DIS*. Because the document description is not yet complete, only the compartments DOC NUMBER, NUMBER OF PAGES, MICROFICHE/OPTICAL DISC and SECURITY LEVEL have been filled in (the first three by the system). He now scans through the document ('browsing') by selecting NEXT PAGE or PRECEDING PAGE. Selecting HARD COPY UNIT tells the COP unit to make a copy. By selecting SCRATCH he can remove the document from the data base. The document description — still incomplete — then disappears from the data base (if he is the only person interested); the document information, however, is retained on the optical disc, but is no longer accessible to him.

The document description can be completed later by the user; he does this by calling up the DOC-TRACK program segment. After he has selected CREATE and then INSERTED? NO, FILED? YES, the system asks him for the document number. When the number has been given, the system checks whether the user has clearance for the document, by consulting the security level entered at an earlier stage. If the answer is yes, the document description (see fig. 21), which is still largely incomplete, appears on his *VDU* screen. He then uses the 'cursor', which indicates a position on a line of the screen, and keys data into the various compartments on the screen. He can enter up to eight keywords in the line KEYWORDS. For our own document file in the Megadoc system, for instance, we have chosen the categories: type, origin, destination, country, organization, project, field of interest and application. When the user has supplied all the data requested, the document description appears again in its final form on the screen. By selecting INSERT? YES he can now transfer the information — still held temporarily in the internal semiconductor memory of the computer — to the external data base located on magnetic discs.

At a later stage the DOC-TRACK program segment can be used to find the document by selecting RETRIEVE. With the aid of the menu the user then supplies the document number, author's name, keyword or date, depending on what he can remember about the particular document. The screen will then show up to four brief descriptions of documents that satisfy the requirements. If there are more than four, a further search can be made by selecting UPWARDS or DOWNWARDS. The document required is then selected by FETCH, which puts the full description on to the *VDU* screen and the first page of the document itself on the document display *DIS*. The user can scan through the document by selecting NEXT PAGE or PRECEDING PAGE. Finally, by selecting HARD COPY UNIT he can obtain a hard copy of the document.

Summary. The Megadoc system consists of a group of peripheral units controlled by a Philips P857 minicomputer and linked together by an electronic switching network. An optical scan converts an A4 document into a quantity of 4×10^6 bits of information, which are then stored in an optical random-access memory, the professional DOR disc. The corresponding document description is recorded in a data base in a conventional magnetic store. A scanning unit generates the binary-digital information of the document within a period of one second and the document display with the associated refresh memory can display the document again in the same time. A storage module for 64 optical discs, which can deliver the desired disc for display within 5 seconds, forms part of an extended modular storage system that can replace many hundreds of cubic metres of more conventional filing. The system also contains units for making hard copies, transmitting documents via public telephone systems and word processing. The software contains programs for storage, distribution, searching and transfer of documents and the recording and selection of document descriptions. The user can communicate with the system by means of terminals consisting of a display unit with keyboard and he can find his way through the programs by answering questions asked by the system.

Fast pressure gauge for hot and corrosive gases

Measuring pressures at high temperatures and in corrosive media is first of all a 'material' problem for the sensors to be used. A number of commercially available pressure gauges employ a metal diaphragm, which senses the pressure as a force acting on the area of the diaphragm. The deformation of this diaphragm, caused by the difference in pressure across it, is measured either capacitively, inductively, resistively or by means of a piezoelectric crystal. Although the measuring range of such gauges is adequate and the output signal is sufficiently linear and reproducible, they are not suitable for use at steady temperatures above 800 K. Their zero drift in the permitted temperature range — arising from thermal expansion — is another difficulty. In addition, the gauges cannot be used in a number of applications because of their limited corrosion resistance, especially at high temperatures.

Because of these various undesirable effects we have developed a new type of gauge for measuring the pressures in high-pressure discharge lamps. *Fig. 1* shows a diagram of the arrangement. The actual output signal *DR* (for 'direct recording') is obtained from the section to the left of the dashed line. The pressure sensor is a modified Bourdon gauge *BM*. As in a conventional Bourdon gauge, the action depends on the microscopic coiling and uncoiling of a spiral tube caused by the pressure difference across the spiral wall. The gauge that we have used (see *fig. 2*) is made of silica glass and has only one turn. This turn consists of a flat silica-glass tube with a wall thickness of only 40 μm. The silica glass is sufficiently resistant to corrosion for our application. Instead of a pointer, a silica-glass plate *RP* provides information about the position of the free end of the spiral.

The angular position of the plate *RP* is measured by a position-sensitive photodiode *PD* (an 'optical potentiometer'), which senses the position of a light beam originating from the laser *L* and reflected by the plate. Electronic circuits ensure that the signal *DR* is not affected by the intensity of the laser beam. The measuring range of the gauge is 60 kPa (reproducibility 100 Pa), corresponding to an angular movement of 3° for the spiral. The maximum operating temperature for the gauge is 1200 K. Because of the lightweight construction of the gauge its resonant frequency is fairly high, about 500 Hz, corresponding to a time constant of the order of 1 ms.

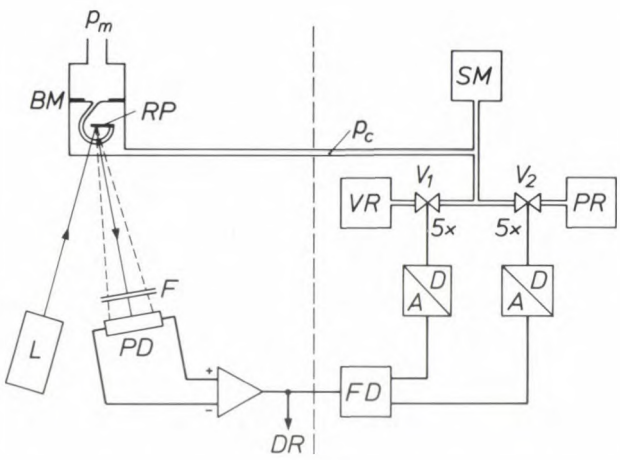


Fig. 1. Basic principle of the pressure measurement. The actual measurement unit is shown to the left of the dashed line. p_m pressure to be measured. *BM* modified Bourdon gauge (see fig. 2). *RP* reflecting silica-glass plate. *L* laser. *F* optical filter. *PD* position-sensitive photodiode. *DR* 'direct-recording' connection. *FD* filter and discriminator. V_1 and V_2 two sets of five parallel valves providing doses in the ratio of 1:2:4:8:16. *VR* vacuum reservoir. *PR* pressure reservoir. *SM* secondary gauge for measuring the compensation pressure p_c.

Fig. 2. The Bourdon gauge used in the instrument. The gauge consists of a spiral with a single turn, made of flat silica-glass tube with a wall thickness of 40 μm. At the end of this spiral is a reflecting silica-glass plate with a matt finish on the reverse side. Because of the lightweight construction the resonant frequency is higher than 500 Hz. On the left-hand side of the gauge there is a window for the incident and reflected laser beams.

As shown in fig. 1, the signal *DR* can also be used as a control signal for compensation of the pressure to be measured p_m. The signal *DR* is amplified, filtered and, depending on the sign, supplied via a first analog-to-digital converter to a set of valves V_1 connected to a vacuum reservoir *VR*, or via a second converter to a set of valves V_2 connected to a pressure reservoir *PR*. In both cases there are five parallel valves providing doses in the ratio of 1:2:4:8:16. In this way the sum of the doses corresponds to the binary signal that opens the associated valves.

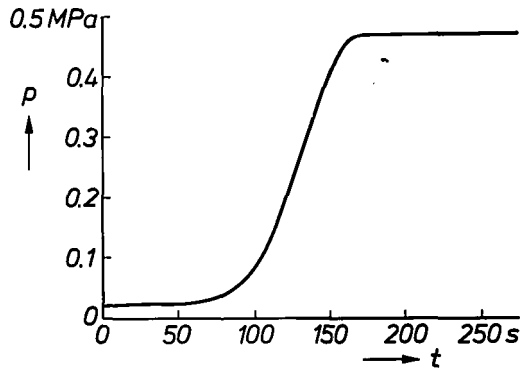


Fig. 3. The pressure-time curve resulting from the evaporation of mercury during the heating-up period in a 400 W metal-halide lamp.

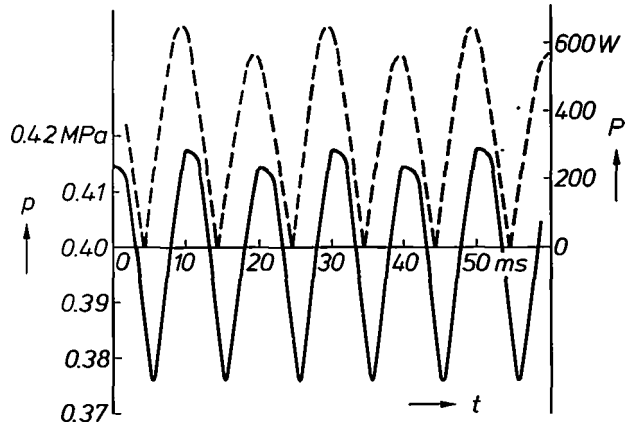


Fig. 4. Periodic pressure variations (*solid curve*) resulting from the 50 Hz periodicity of the electrical mains supply in a 400 W metal-halide lamp (*dashed curve*).

This control scheme enables a compensation pressure p_c to be produced that has the same magnitude as the pressure to be measured p_m. The pressure p_c is measured by a secondary gauge *SM*. Because we are concerned here with pressure measurement of a non-corrosive gas at room temperature the gauge *SM* does not have to meet any special requirements.

We stated earlier that the time constant of the gauge *BM* is only about 1 ms. Because the pressure compensation also takes place relatively quickly — within approximately 50 ms — the time constant of the gauge *SM* used in our measurements determines the speed of the measurement: approximately 300 ms.

Fig. 3 illustrates a pressure curve measured during the heating-up period of a 400 W metal-halide lamp. The increase in the pressure is mainly determined by the evaporation of mercury. The rapid response of the actual gauge *BM* can clearly be seen in *fig. 4*, which illustrates the periodic variations in pressure in the same lamp caused by the 50 Hz periodicity of the electrical mains supply. During these measurements the mean pressure of 0.4 MPa is compensated by the control system, while the signal due to the pressure variations is measured through the 'direct-recording' connection *DR* and thus originates from the gauge *BM*.

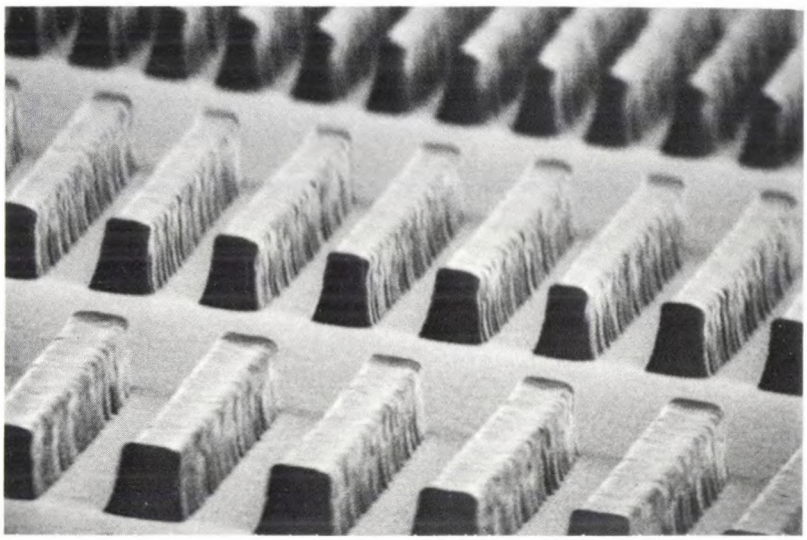
T. Zaengel
G. A. Lens

Dr T. Zaengel and G. A. Lens are with Philips GmbH Forschungslaboratorium Aachen, Aachen, West Germany.

Negative electron resists for VLSI

J. C. Jagt and P. W. Whipps

Integrated-circuit technology now stands at the start of a new era, in which VLSI will soon encompass the production of circuits with more than 100 000 components on a chip. To make this a practical possibility a great deal of research is now under way to bring methods and materials to perfection. Electron lithography requires suitable negative resists, and some of our work is described in this article.



Introduction

In the production of integrated circuits [1] lithography is used as a method for preparing patterns, which are transferred to a substrate by means of etching processes (*fig. 1*). The procedure is normally as follows. First of all a layer of resist is exposed through a mask of the required pattern to radiation (e.g. light) to which it is sensitive. Next the resist is treated with a developer, which dissolves either the exposed or the unexposed areas of the resist — in the first case the resist is said to be *positive*, in the second case *negative*. Finally the unprotected parts of the underlying layer are subjected to etching processes.

In fig. 1, apertures are etched in a masking layer of silicon dioxide; when this has been done the silicon can be doped appropriately. The conducting strips between the doped areas are formed by etching away an evaporated layer of aluminium. A similar method is applied to prepare the required mask by etching away chromium evaporated on glass.

Dr J. C. Jagt is with Philips Research Laboratories, Eindhoven; P. W. Whipps, M.A., is with Philips Research Laboratories (PRL), Redhill, Surrey, England.

If light is used — still the simplest method — the wavelength of the light sets a fundamental limit to the size of the details in the pattern, since diffraction effects cause blurring of details smaller than 1 μm. At present, photolithography is the most widely used method for the preparation of integrated circuits but the lower limit to the dimensions of particular features necessarily corresponds to an upper limit to the number of components on a chip. In the future there will undoubtedly be a need for further miniaturization and more complex circuits in chips containing more than 10^5 components (VLSI), and so various other lithographic methods are being developed in the meantime. These include electron, X-ray and ion lithography [2], which should give higher resolution.

At present the most promising of these methods for VLSI technology seems to be electron lithography, particularly the method in which the pattern is produced by means of an extremely fine computer-controlled beam that 'writes' on the resist layer. The electron-beam pattern generator developed for this purpose has been described in an earlier article in this

journal [3]. The generator is already in extensive use for mask production in LSI manufacture. In this application its resolution may be better than is strictly necessary, and it is used mainly because the time from design to completed mask is much shorter than with the conventional optical methods.

In future VLSI technology the electron-beam pattern generator will be used not only for mask production but also for writing directly on the chip. In 'direct writing' it will also be necessary to switch from the conventional wet-etching procedures to dry-etching methods. Wet etching can lead to severe undercutting with loss of small detail.

The most common dry-etching methods are plasma etching, in with the solid material is converted into volatile products by reactive particles in a gas discharge [4], and 'sputter etching', where the solid substance is physically removed by bombardment with noble-gas ions [5]; a combination of both, reactive ion etching, is also used.

In this article we deal with a number of recently developed *negative* electron resists that are particularly resistant to dry etching and which therefore seem suitable for application in VLSI technology.

A general introduction to the nature and operation of both positive and negative electron resists has been given in an earlier article in this journal [6]. This earlier article may be consulted for further data on *positive* electron resists, which also have potential uses in VLSI technology. It is obviously desirable to have both positive and negative electron resists available for VLSI production, so that we can choose the resist that gives the shortest 'writing time' for any particular pattern.

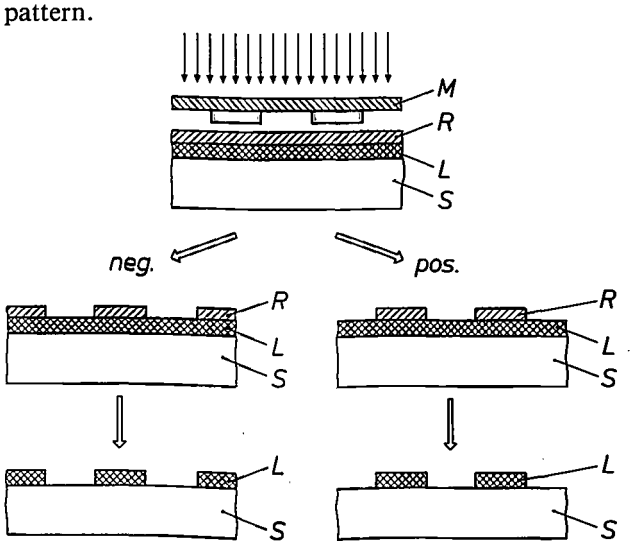


Fig. 1. The method of forming patterns in integrated-circuit production. *S* silicon wafer. *L* masking layer. *R* resist layer. *M* mask. After irradiation and development of the resist layer the pattern on the left or the right is obtained, depending on whether a negative or a positive resist is used. After the unprotected layer underneath has been etched away and the rest of the layer of resist removed, the appropriate dopants can be introduced.

Sensitivity, contrast and swelling

We have already mentioned two of the criteria used in assessing the suitability of an electron resist for VLSI — the resolution and the resistivity to dry etching. Other criteria are the *sensitivity*, the *contrast* and the degree of *swelling* during development. The last two factors also affect the resolution.

Sensitivity

An electron resist consists mainly of linear molecules of an organic polymer. As the energy of the electrons bombarding a resist is several times greater (e.g. 20 keV) than the energy of the bonds in the polymer molecule, a great variety of reactions can occur as a result of irradiation. The nature of the action of the resist — i.e. whether it is positive or negative — depends on what happens when the polymer molecule is irradiated with high-energy electrons. If it breaks down into smaller molecules ('scission'), then the action is positive. If it combines to form a three-dimensional network that is insoluble in the developer (cross-linking), then the action is negative [6]. What happens generally depends on the main polymer chain — the presence of tertiary carbon atoms in the main chain usually leads to cross-linking, for example — but it can also be determined by the nature of the side groups. By introducing reactive groups such as alkenyl and epoxy groups (see fig. 5*a*) into the polymer, negative resists of very high sensitivity can be obtained.

The sensitivity of a negative resist is defined as the 'dose' of electrons necessary to convert unit area of the resist into an insoluble layer, whose thickness is measured as a fraction of the original thickness after development. In practice it is usual to determine the dose $D_g^{0.5}$ at which 50% of the original layer thickness remains after development, from the characteristic or exposure curve (*fig. 2*). The extrapolated values D_e^0 and $D_e^{1.0}$ can also be estimated.

The sensitivity depends not only on the chemical structure of the polymer, but also on the molecular weight. As the constituent molecules of a negative

[1] For a more detailed description, see A. Schmitz, Philips tech. Rev. 27, 192, 1966, and H. Bosma and W. G. Gelling, Philips tech. Rev. 37, 267, 1977, in particular p. 270.

[2] For electron lithography see for example M. J. Bowden, CRC crit. Rev. Solid State and Mat. Sci. 8, 223, 1978/79, and R. P. Kramer, Ned. T. Natuurk. A 46, 4, 1980 (in Dutch); for X-ray lithography see for example E. Spiller and R. Feder, in: H.-J. Queisser (ed.), X-ray optics (Topics in appl. Phys. 22), p. 35, Springer, Berlin 1977; for ion lithography see for example R. L. Seliger and P. A. Sullivan, Electronics 53, 27 March 1980, 142.

[3] J. P. Beasley and D. G. Squire, Philips tech. Rev. 37, 334, 1977.

[4] See for example H. Kalter and E. P. G. T. van de Ven, Philips tech. Rev. 38, 200, 1978/79.

[5] See for example H. Dimigen and H. Lüthje, Philips tech. Rev. 35, 199, 1975.

[6] E. D. Roberts, Philips tech. Rev. 35, 41, 1975.

electron resist increase in size, the dose of radiation necessary to produce a sufficiently extensive three-dimensional network for the polymer to become insoluble in the developer decreases, so that the sensitivity of the resist increases.

The degree of sensitivity required in the production process depends on the method used and the apparatus. When the electron-beam pattern generator is used, the complete pattern is written in a sequence of small steps and the sensitivity of the resist must be high if the writing time is to be kept short. The present requirements for resist sensitivity (which are based on the writing speed and beam current of the present generation of pattern generators) are about 5 to 10 $\mu C/cm^2$.

Contrast

During the irradiation electrons are scattered by the resist. These electrons — and especially the electrons scattered back into the layer of resist — cause a broadening of the region in which energy absorption takes place, making it larger than the nominally 'exposed' areas, and hence limiting the resolution (*fig. 3*). To minimize the effect of this back-scatter it is important that the resist should have a high contrast, as defined from the slope of the characteristic curve in fig. 2: $\gamma = 1/\log(D_e^{1.0}/D_e^0)$. The greater the slope, the less the likelihood that back-scattered electrons will leave an insoluble resist layer in unexposed areas. The contrast γ of an electron resist should usually be at least 1.5; a contrast of 2.0 is considered to be good.

The contrast depends partly on the spread in the size of the polymer molecules in the resist. The irradiation doses are always kept to the minimum necessary to make the resist insoluble; these doses depend on the average size of the polymer molecules. Consequently, in a 'monodisperse' resist (i.e. one with a small spread in the magnitude of the polymer molecules) the small doses of back-scattered electrons will not affect the unexposed resist sufficiently to make it insoluble. However, as the polydispersity of the resist increases there are an increasing number of polymer molecules with high molecular weight that will be affected by low levels of exposure (and are therefore also likely to be cross-linked by back-scattered electrons). Similarly, any molecules of low molecular weight require a higher exposure level for cross-linking. The overall effect is that the exposure curve (as in fig. 2) is broadened, becoming less steep.

Swelling

During development the non-irradiated areas of the resist are dissolved in the developer. Since the resist is always irradiated with the minimum dose, as we saw above, the degree of cross-linking in the polymer is rather low and it is relatively easy for solvent molecules (from the developer) to be absorbed into the network formed by the irradiated resist, giving rise to swelling. This effect is obviously more serious when high resolutions are required (*fig. 4*).

The amount of swelling in an irradiated resist depends not only on the cross-linking of the network

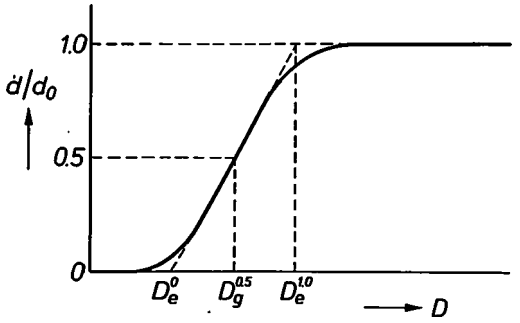


Fig. 2. Schematic plot of the characteristic curve of a negative electron resist. The thickness d of the resist after development, expressed as a fraction of the original thickness d_0, is shown as a function of the exposure dose D (logarithmic scale). The sensitivity of the resist is often quoted in practice in terms of the dose $D_g^{0.5}$ at which a half of the original thickness remains after development. For theoretical work the extrapolated values D_e^0 and $D_e^{1.0}$ are often used, at which the remaining relative thicknesses are 0 and 1 respectively. The slope of the straight part of the curve is taken as a measure of the contrast: $\gamma = 1/\log(D_e^{1.0}/D_e^0)$.

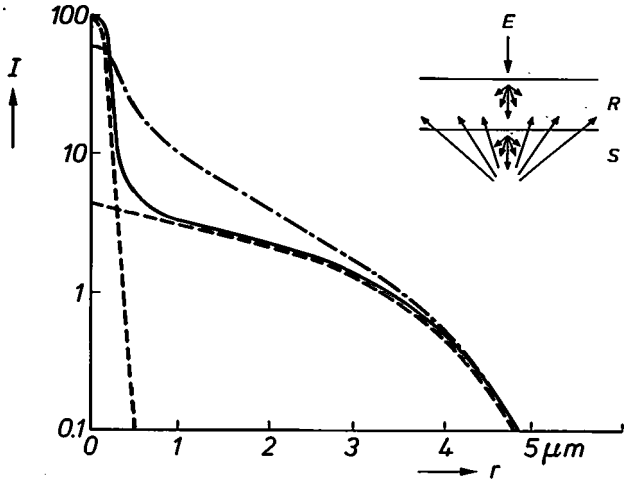


Fig. 3. The solid curve gives a schematic representation of the distribution of electrons at the surface of a layer of resist R, for writing with a gaussian electron beam E (20 keV): the curve shows the radiation intensity I plotted against the distance r from the centre of the beam. The dashed lines show the separate contributions from the incident and back-scattered electrons. The back-scattering causes a broadening of the area in which energy is absorbed. The chain-dotted line indicates the electron distribution at the boundary surface of the resist layer (about 1 μm thick) and the substrate S. It is clear that the forward scattering also contributes to the broadening. The forward- and back-scattered electrons limit the resolution, particularly when closely adjacent lines have to be produced ('proximity effect'), since the intensity distributions of the separate lines then overlap considerably. (From T. R. Neill and C. J. Bull, in: Microcircuit Engineering 80, Proc. Int. Conf. on Microlithography, Amsterdam 1980, fig. 1 on page 51.)

produced but also on its flexibility and on the strength of the developer. To suppress the swelling, the degree of cross-linking and the rigidity of the network should be increased as far as possible and the developer should be so weak that it is only just capable of dissolving the non-irradiated resist.

A low contrast often accentuates the tendency to swelling, because a network with a low degree of cross-linking is then formed at the edges of the pattern the developer acts on.

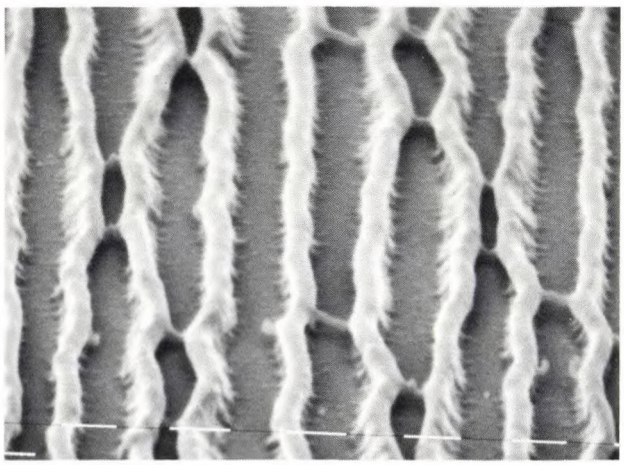


Fig. 4. Defects caused by swelling in a pattern drawn in resist containing epoxy groups that has been bombarded with electrons (25 keV) and then developed (SEM photograph: the marker is 1 μm). The materials and conditions were specially selected to give an exaggerated demonstration of the effect.

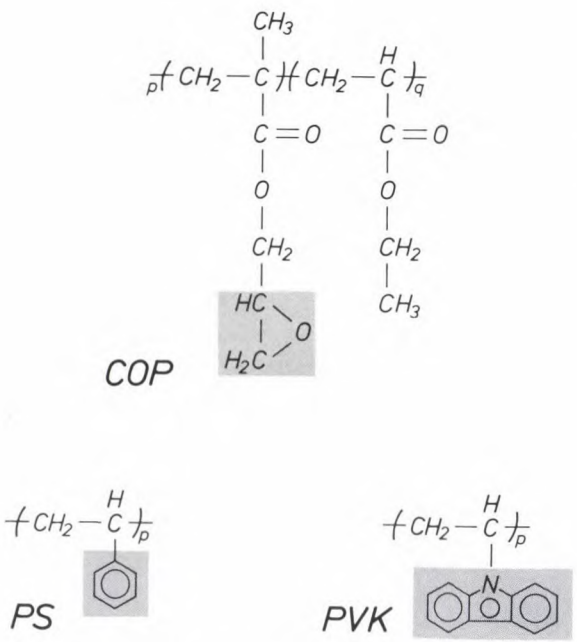


Fig. 5. Structural formulae of three negative electron resists referred to in the article: *a*) poly(glycidyl methacrylate-co-ethyl acrylate) or COP, *b*) polystyrene (PS) and *c*) polyvinyl carbazole (PVK). While the first material has a high sensitivity, largely because of its epoxy groups (*grey background*), the other two are outstanding for their high dry-etch resistance, due respectively to their benzene and carbazole groups (*also with a grey background*).

Research strategy

In the search for suitable negative electron resists for VLSI, many laboratories initially investigated the potential usefulness of commercial poly(glycidyl methacrylate-co-ethyl acrylate) (COP), which is widely used in mask manufacture; see *fig. 5*. This material has a high electron sensitivity (about 1 $\mu C/cm^2$), mainly because of the presence of the reactive epoxy groups. However, it gives a rather low contrast (about 1) and its resistance to dry-etching is relatively poor. A material such as this also swells appreciably when it is developed. Attempts have been made at several laboratories to improve these properties, while maintaining its high sensitivity, by introducing variations into the structure.

In our investigations we adopted a different strategy for our research, based on two other materials: polystyrene (PS) and the structurally related polyvinyl carbazole (PVK); see fig. 5. A characteristic of both polymers is the presence of large numbers of aromatic groups: benzene rings in polystyrene and carbazole groups in polyvinyl carbazole. The presence in the polymer of such aromatic groups increases the dry-etch resistance. At the same time these aromatic groups reduce the sensitivity of the material to cross-linking. Polystyrene does have good plasma-etch resistance and low sensitivity. Nevertheless, because of its excellent contrast this polymer has been used in several laboratories as a starting material in efforts to produce a suitable electron-sensitive resist.

We shall now present some results from a number of investigations carried out at Philips Research Laboratories in Eindhoven and in Redhill. In our efforts to obtain a higher sensitivity with the two polymers mentioned above there were four different approaches. We examined the formation of copolymers containing epoxy groups, the effects of increasing the molecular weight, the introduction of halogen substituents, and the use of a cross-linking agent.

Experimental results

Copolymerization

The properties of two or more polymeric materials can be combined by means of a method known as 'copolymerization', in which a polymer is synthesized from two (or more) different monomers. The basic monomers we used were styrene and glycidyl methacrylate, which contains the epoxy group (*fig. 6*).

The copolymer obtained in this way proved in fact to have a substantially better sensitivity than polystyrene. The introduction of the glycidyl-methacrylate groups, however, causes significant deterioration of the plasma-etch resistance compared with that of

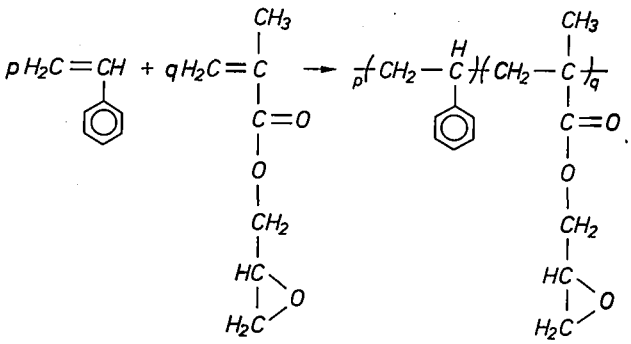


Fig. 6. Structural formula of a polymer synthesized from the monomers styrene (*the molecule at the upper left*) and glycidyl methacrylate. The objective in synthesizing this polymer was to obtain a resist combining the good dry-etch resistance of polystyrene with the high sensitivity of poly(glycidyl methacrylate).

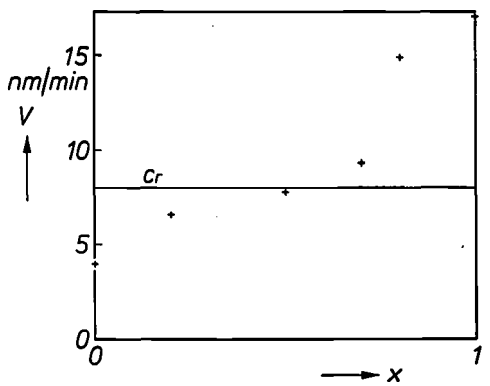


Fig. 7. Plasma-etch resistance of various copolymers of styrene with glycidyl methacrylate (+), compared with that of chromium (*solid line*). The proportion of the glycidyl-methacrylate molecules and styrene in the copolymer is plotted along the *x*-axis; the rate *V* of the etching process in a mixture of chlorine, helium, and oxygen in the ratio of 20:78:2 at a pressure of 67 Pa (0.5 torr) is plotted along the *y*-axis.

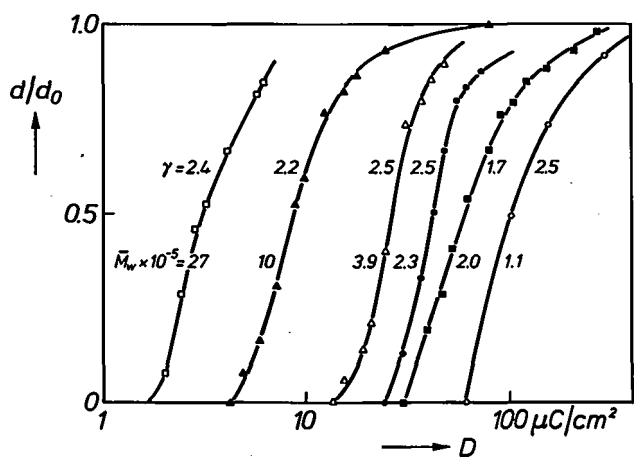


Fig. 8. Characteristic curves at 25 kV for monodisperse polystyrene samples of different average molecular weight $\bar{M}_w$ [7]. The various contrast factors γ and a characteristic curve (■) for a polydisperse polystyrene sample of average molecular weight of 2.0×10^5 are also included.

polystyrene (*fig. 7*). The contrast of this copolymer is also worse than that of polystyrene. Any gain in sensitivity is thus lost on the other properties. We therefore looked for a different approach to the desired improvement in sensitivity without impairing the other properties. One such approach could be to increase the molecular weight.

Use of higher molecular weights

Figs 8 and *9* show how the sensitivities of monodisperse polystyrene samples increase with molecular weight, without any loss in contrast. The contrast of a polydisperse sample is also seen to be sufficiently high. Similar results have been obtained with polyvinyl carbazole (fig. 9).

It may be deduced from fig. 8 that an average molecular weight $\bar{M}_w$ [7] of about 2×10^6 is required for a sensitivity $D_g^{0.5}$ of 5 μC/cm², a desirable value for the manufacture of masks. For direct irradiation of the silicon wafer even higher sensitivities will probably be needed, implying molecular weights greater than 2×10^6. At such high molecular weights, however, the resist solution becomes difficult to handle because of its high viscosity. Although promising results were obtained from this approach we felt that it would be more practical to try to find a resist with a rather lower molecular weight.

Introduction of chlorine and bromine

It was known from earlier work that the substitution of chlorine and bromine in the benzene rings of polystyrene increases the sensitivity to gamma rays [8]. In our work the sensitivity to electrons was also found to be markedly improved compared with unsubstituted polystyrene.

The halogen atoms can be introduced in two ways: before polymerization, in the monomer, or after polymerization, in the polymer. Halogen substitution after the polymerization seemed less advantageous to us, since it often has the effect of impairing and damaging the polymer chains, and this leads to

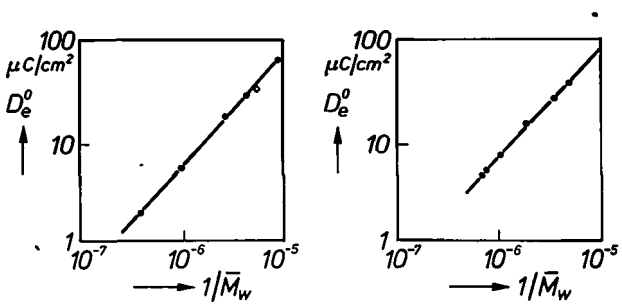


Fig. 9. *Left:* Sensitivity D_e^0 at 25 kV of monodisperse polystyrene as a function of average molecular weight $\bar{M}_w$. The sensitivity of a polydisperse polystyrene sample is also shown (○). *Right:* Sensitivity D_e^0 of polyvinyl carbazole samples.

broadening of the molecular-weight distribution and hence to reduced contrast. The approach adopted was therefore to introduce the halogen substituent in the monomer stage. Polymers were formed from *p*-chlorostyrene, and copolymers were also prepared of this monomer with styrene. The best sensitivity results from a copolymer containing 20 moles per cent of *p*-chlorostyrene (*fig. 10*): $D_g^{0.5}$ for material of average molecular weight $\bar{M}_w = 8 \times 10^4$ is about 4.5 μC/cm² compared with 115 μC/cm² for unsubstituted polystyrene of the same average molecular weight. The contrast value for this material and its dry-etch resistance (see *fig. 11*) were also very good. *Fig. 12* shows an SEM photograph of a pattern written in this material.

The sensitivity of polyvinyl carbazole is also greatly improved by the introduction of halogen atoms. For this polymer the introduction of *bromine* atoms proved to be the easiest procedure. In this case the highest sensitivity, 1.3 μC/cm², was found in the polymer prepared from the monomer 3-bromo-N-vinyl carbazole. The contrast of the brominated polyvinyl carbazole, however, is generally somewhat lower than that of the chlorostyrene copolymer. The definition of the edges of the patterns is also rather worse.

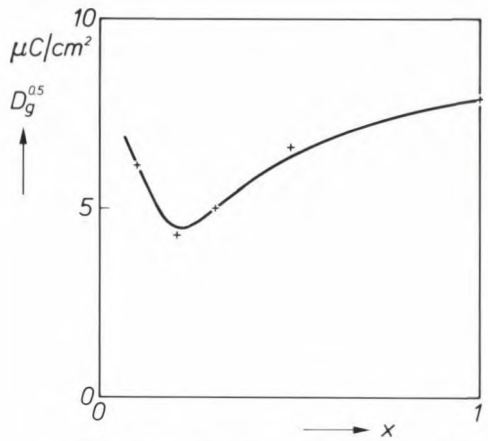


Fig. 10. Sensitivity $D_g^{0.5}$ at 20 kV of samples of styrene/*p*-chlorostyrene copolymers with molecular weights $\bar{M}_w$ of approximately 8×10^4; *x* indicates the molecular proportion of chlorostyrene. The copolymerization allows a sensitivity of 4.5 μC/cm² to be achieved.

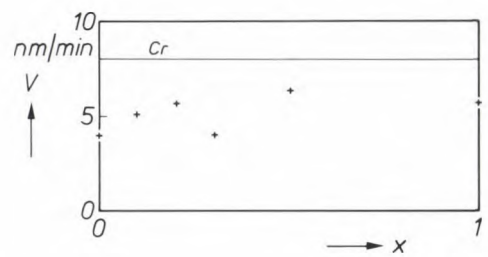


Fig. 11. Plasma-etch resistance for styrene/*p*-chlorostyrene copolymers; *x* is as in fig. 10, and etch conditions are as described for fig. 7, again including comparison with chromium.

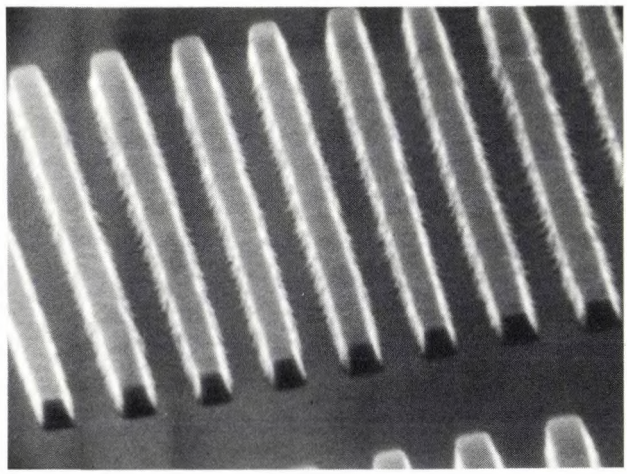


Fig. 12. SEM photograph of a pattern of lines made with poly(styrene-co-*p*-chlorostyrene), with a width and gap of 0.7 μm.

Use of cross-linking agents

Substances or additives known as cross-linking agents are widely used in photolithography to obtain resists of high sensitivity. The best known of these are the 'bisazides'. These compounds break down when they are irradiated, nitrogen being split off and a bridge being formed between two polymer chains (*fig. 13*).

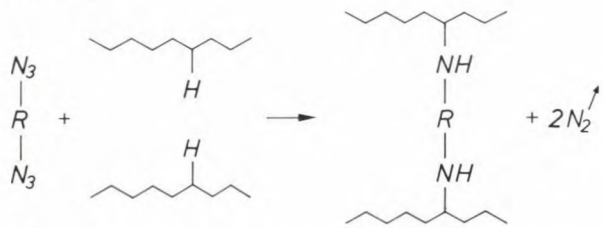


Fig. 13. Cross-linking of polymers by a bisazide (N_3RN_3).

We found that the addition of small amounts of bisazide to polystyrene increased the sensitivity to electrons by a factor of 10 and also increased the contrast. *Fig. 14* and *Table I* present the results for polystyrenes and polyvinyl carbazoles mixed in this way with the bisazide N_3-⟨⟩-CH=CH-⟨⟩-N_3. Their sensitivities are all better than 7 μC/cm², their contrast is improved, and their dry-etch resistance and swelling behaviour

[7] The average molecular weight of a polymer may be expressed as a 'number-averaged' molecular weight $\bar{M}_n = \frac{\Sigma n_i M_i}{\Sigma n_i}$, where n_i is the number of molecules with a molecular weight M_i in the *i*th fraction, or as a 'weight-averaged' molecular weight $\bar{M}_w = \frac{\Sigma w_i M_i}{\Sigma w_i}$, where w_i is the weight of the molecules of molecular weight M_i in the *i*th fraction. The ratio $\bar{M}_w/\bar{M}_n$ (the polydispersity) is a measure of the width of the molecular-weight distribution. Ideal monodisperse polymers have $\bar{M}_w/\bar{M}_n = 1$, while practical polymers have this ratio >1. The sensitivity of a resist, measured as D_e^0 (fig. 2), is inversely proportional to $\bar{M}_w$.

[8] W. Burlant, J. Neerman and V. Serment, J. Polymer Sci. **58**, 491, 1962.

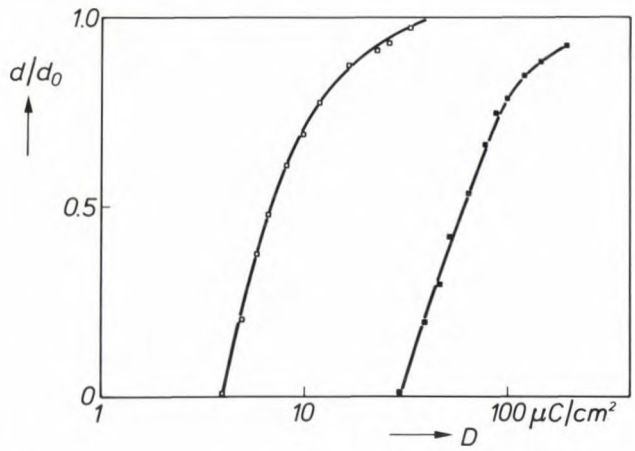


Fig. 14. Characteristic curve at 25 kV (■) compared with the same material mixed with the bisazide 4,4′-diazidostilbene: N_3-⟨⟩-CH=CH-⟨⟩-N_3. Similar results have been obtained with the bisazide 4,4′-*bis*(azidosulphonyl) biphenyl: N_3SO_2-⟨⟩-⟨⟩-SO_2N_3. The average molecular weight $\bar{M}_w$ of the polystyrene molecules was 2.0×10^5, while the spread in molecular weight is indicated by $\bar{M}_w/\bar{M}_n = 2.95$ [7].

Table I. Sensitivity $D_g^{0.5}$ (at 25 kV), contrast γ, average molecular weight $\bar{M}_w$ and the ratio $\bar{M}_w/\bar{M}_n$ [7] of polystyrene (PS) and polyvinyl carbazole (PVK) samples, with and without the bisazide B:N_3-⟨⟩-CH=CH-⟨⟩-N_3. The addition of the bisazide can be seen to be beneficial in all cases.

Resist	$\bar{M}_w \times 10^{-5}$	$\bar{M}_w/\bar{M}_n$	$D_g^{0.5}$	γ
PS	2.0	2.95	60	1.7
PS + B	2.0	2.95	7.3	2.1
PS	3.9	1.0	27	2.5
PS + B	3.9	1.0	2.8	3.0
PS	10.2	1.15	9	2.2
PS + B	10.2	1.15	0.8	3.0
PVK	13.1	1.8	8.1	2.1
PVK + B	13.1	1.8	2.2	2.2

are not inferior to those of pure polystyrene or polyvinyl carbazole.

Good results were also found in later experiments when the bisazide N_3SO_2-⟨⟩-⟨⟩-SO_2N_3 was added (*fig. 15*). This bisazide is easier to prepare than the bisazide used first. Also, its sensitivity to visible light is negligible, unlike the bisazide mentioned above, which has the disadvantage of reacting both to electrons and to light. The title photograph shows a pattern with gaps of 0.6 μm and 0.8 μm that was made with the aid of a polystyrene resist containing this bisazide. After the resist had been developed the polysilicon substrate was removed at the exposed locations by plasma-etching.

Our provisional conclusion is that, of the various materials we have investigated, the following three seem to be particularly suitable for use as negative electron resists in LSI technology and also for future VLSI technology: a copolymer of styrene with *p*-chlorostyrene, polystyrene with a bisazide cross-linking agent, and polyvinyl carbazole with bisazide. These three resists are now being tested with an electron-beam pattern generator for their practical usefulness, and the first results are promising.

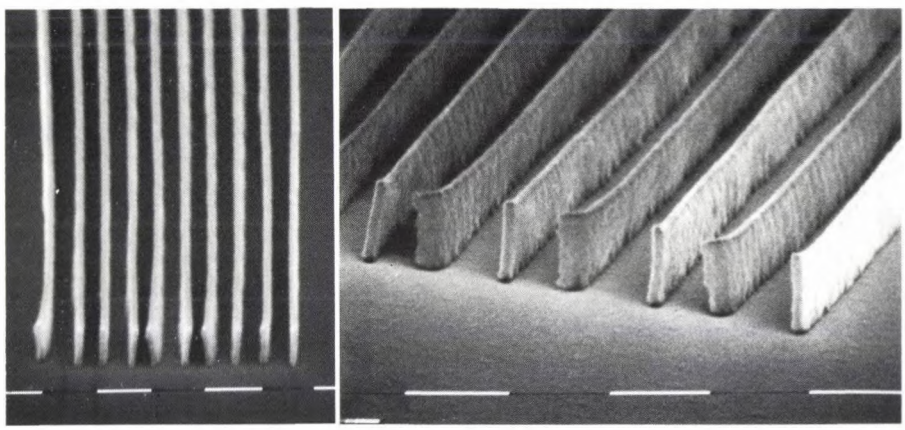


Fig. 15. SEM photograph of patterns made with polyvinyl carbazole and the bisazide 4,4′-*bis*-(azidosulphonyl)biphenyl. (The marker divisions indicate 1 μm.)

Summary. The masks used in the production of LSI circuits are now increasingly manufactured with the aid of an electron-beam pattern generator. If this method is also to be used for VLSI, electron resists will be required with a sensitivity of 5 to 10 μC/cm², a contrast factor greater than 1.5 and the ability to withstand the dry-etch processes that will replace the wet methods now used. At the Philips Research Laboratories in Eindhoven and Redhill, England, efforts have been made to synthesize negative electron resists that are more resistant to dry etching than poly(glycidyl methacrylate-co-ethyl acrylate) (COP) and have a higher electron sensitivity than polystyrene. The methods used were copolymerization of glycidyl methacrylate with styrene, the use of higher molecular weights of polystyrene and of the related polyvinyl carbazole, halogen substitution in these materials, and the use of cross-linking agents. The most suitable electron resists seem at present to be a copolymer containing chlorostyrene, and mixtures of polystyrene or polyvinyl carbazole with a bisazide as cross-linking agent.

Scientific publications

These publications are contributed by staff of laboratories and plants that form part of or cooperate with enterprises of the Philips group of companies, particularly by staff of the following research laboratories:

Philips Research Laboratories, Eindhoven, The Netherlands	*E*
Philips Research Laboratories, Redhill, Surrey RH1 5HA, England	*R*
Laboratoires d'Electronique et de Physique Appliquée, 3 avenue Descartes, 94450 Limeil-Brévannes, France	*L*
Philips GmbH Forschungslaboratorium Aachen, Weißhausstraße, 51 Aachen, Germany	*A*
Philips GmbH Forschungslaboratorium Hamburg, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany	*H*
Philips Research Laboratory Brussels, 2 avenue Van Becelaere, 1170 Brussels (Boitsfort), Belgium	*B*
Philips Laboratories, N.A.P.C., 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A.	*N*

G. Aaron, S. Boronkay & J.-J. Poubeau: La télécopie: joindre le dessin à la parole.
La Recherche **9**, 1032-1034, 1978 (No. 94). *L*

W. J. Bartels & W. Nijman: X-ray double-crystal diffractometry of $Ga_{1-x}Al_xAs$ epitaxial layers.
J. Crystal Growth **44**, 518-525, 1978 (No. 5). *E*

P. Baudet, M. Binet, D. Boccon-Gibod & L. Hollan: Self-aligned GaAs power MESFETs for X-band.
8th Eur. Microwave Conf., Paris 1978, pp. 391-395. *L*

V. Belevitch: On the convergence of infinite resistance ladders.
Circuit theory and design, ed. G. S. Moschytz & J. Neirynck, pp. 29-45; Georgi, St Saphorin 1978. *B*

P. Blood, W. L. Brown* & G. L. Miller* (* Bell Laboratories, Murray Hill, N.J.): Incorporation of implanted In and Sb in silicon during amorphous layer regrowth.
J. appl. Phys. **50**, 173-182, 1979 (No. 1). *R*

D. Boccon-Gibod & P. Harrop: High performance GaAs Schottky barrier diodes using a cantilevered metal contact.
8th Eur. Microwave Conf., Paris 1978, pp. 696-700. *L*

H. Bouma (Institute for Perception Research, Eindhoven): Visual search and reading: eye movements and functional visual field: a tutorial review.
Attention and performance VII, Proc. 7th Int. Symp., Sénanque 1976, pp. 115-147; 1978.

P. W. J. M. Boumans: A tutorial review of some elementary concepts in the statistical evaluation of trace element measurements.
Spectrochim. Acta **33B**, 625-634, 1978 (No. 9). *E*

P. W. J. M. Boumans: Inductively coupled plasma-atomic emission spectroscopy: its present and future position in analytical chemistry.
Óptica pura y apl. **11**, 143-171, 1978 (No. 3). *E*

D. G. Bouwhuis (Institute for Perception Research, Eindhoven): Visual recognition of words.
Thesis, Nijmegen 1979.

R. J. Brewer: A low noise CCD output amplifier.
Tech. Dig. 1978 Int. Electron Devices Meeting, Washington D.C., pp. 610-612. *R*

F. J. A. den Broeder: Surface diffusion of tungsten and tungsten-rhenium alloys.
High Temp. – high Press. **10**, 131-133, 1978 (No. 2). *E*

S. D. Brotherton & A. Gill: The energy level of thallium in silicon.
Appl. Phys. Letters **33**, 953-955, 1978 (No. 11). *R*

J. Bruinink & C. G. A. Kregting: The voltammetric behavior of some viologens at SnO_2 electrodes.
J. Electrochem. Soc. **125**, 1397-1401, 1978 (No. 9). *E*

C. Bull, P. Ashburn, G. R. Booker (all with University of Oxford) **& K. H. Nicholas:** Effects of dislocations in silicon transistors with implanted emitters.
Solid-State Electronics **22**, 95-104, 1979 (No. 1). *R*

K. H. J. Buschow, A. M. van Diepen, N. M. Beekmans & J. W. M. Biesterbos: Charge transfer and magnetic properties in amorphous alloys of Fe with Y, Th or Zr.
Solid State Comm. **28**, 181-185, 1978 (No. 2). *E*

F. M. A. Carpay & S. K. Kurtz: A multiple-lognormal model of normal grain growth.
Processing of crystalline ceramics, ed. Hayne Palmour III, R. F. Davis & T. M. Hare, pp. 217-229; Plenum Press, New York 1978. *E*

M. Cathelin, J. P. Chané, G. Durand & M. Gavant: Microwave frequency digital circuits using GaAs MESFETs with a planar self-aligned technology.
8th Eur. Microwave Conf., Paris 1978, pp. 55-60. *L*

M. M. Choy, W. K. Zwicker & S. R. Chinn (M.I.T., Lexington, Mass.): Emission cross section and flash-lamp-excited NdP_5O_{14} laser at 1.32 μm.
Appl. Phys. Letters **34**, 387-388, 1979 (No. 6). *N*

W. J. Dallas, R. Linde & H. Weiss: Simplified production of spatial inverse filters.
Optics Letters **3**, 247-249, 1978 (No. 6). *H*

P. Delsarte, Y. Genin & Y. Kamp: A criterion for two-variable positivity and boundedness.
Circuit theory and design, ed. G. S. Moschytz & J. Neirynck, pp. 365-370; Georgi, St Saphorin 1978. *B*

R. Dessert, P. Harrop, B. Kramer & T. Vlek (Philips Video Division, Eindhoven): All FET front-end for 12 GHz satellite broadcasting reception.
8th Eur. Microwave Conf., Paris 1978, pp. 638-644. *L*

P. A. Devijver: Nonparametric estimation by the method of ordered nearest neighbor sample sets.
Proc. 4th Int. Joint Conf. on Pattern recognition, Kyoto 1978, pp. 217-223. *B*

A. M. van Diepen & F. K. Lotgering: Preparation and Mössbauer spectra of CoV_2O_4 spinel doped with Fe^{2+} on tetrahedral sites.
Solid State Comm. **28**, 951-955, 1978 (No. 11). *E*

E. W. Dijkstra (Burroughs Corporation, Eindhoven), **L. Lamport** (SRI International, Menlo Park, Calif.), **A. J. Martin, C. S. Scholten & E. F. M. Steffens:** On-the-fly garbage collection: an exercise in cooperation.
Comm. ACM **21**, 966-975, 1978 (No. 11). *E*

A. G. Dirks & H. J. Leamy (Bell Laboratories, Murray Hill, N.J.): Magnetic and structural properties of some amorphous rare earth - transition metal thin films.
IEEE Trans. **MAG-14**, 835-837, 1978 (No. 5). *E*

J. M. L. Engels: The drift velocity of excess electrons in fluid methane, argon and mixtures of methane and argon.
Thesis, Delft 1979. *E*

D. den Engelsen, A. A. C. M. van Gorp & L. Heyne: Potentiostatic transients at rotating solid-electrolyte electrodes.
J. Electrochem. Soc. **126**, 242-248, 1979 (No. 2). *E*

G. Engelsma: Inhibition of phenylalanine ammonia-lyase by cinnamic acid derivatives.
Regulation of secondary product and plant hormone metabolism (12th FEBS Meeting, Dresden 1978, Vol. 55), ed. M. Luckner & K. Schreiber, pp. 163-172; Pergamon Press, Oxford 1979. *E*

K. Enke & W. Tolksdorf: Continuously recording refractive index spectrograph for transparent and opaque insulators and semiconductors.
Rev. sci. Instr. **49**, 1625-1628, 1978 (No. 12). *H*

V. Evers & P. M. van den Avoort: Research on new media for home-based learning.
1978 Frontiers in Education Conf., Lake Buena Vista (Florida), pp. 260-267. *E*

G. Frank, H. Köstlin & A. Rabenau (Max-Planck-Institut für Festkörperforschung, Stuttgart): X-ray and optical measurements in the In_2O_3-SnO_2 system.
Phys. Stat. sol. (a) **52**, 231-238, 1979 (No. 1). *A*

P. E. C. Franken: The influence of the grain boundary on the temperature coefficient of Ti-substituted telecommunication ferrites.
IEEE Trans. **MAG-14**, 898-899, 1978 (No. 5). *E*

G. Frens: On coagulation in the primary minimum.
Faraday Disc. Chem. Soc. No. 65, 146-155, 1978. *E*

R. G. Gossink: SIMS analysis of a field-assisted glass-to-metal seal.
J. Amer. Ceramic Soc. **61**, 539-540, 1978 (No. 11/12). *E*

H. C. de Graaff & J. G. de Groot: The SIS tunnel emitter.
Tech. Dig. 1978 Int. Electron Devices Meeting, Washington D.C., pp. 333-335. *E*

J. Haisma, E. Hugues* & C. Babolat* (* Centre de Recherches et de Calculs Optiques, Courbevoie): Realization of a bi-aspherical objective lens for the Philips Video Long Play system.
Optics Letters **4**, 70-72, 1979 (No. 2). *E*

J. P. Hazan: Applications of Titus and Phototitus to data processing.
Proc. 1978 Int. Optical Computing Conf., London, pp. 17-20. *L*

H. Heitmann, P. Hansen, K. Witter & R. Spohr (Ges. f. Schwerionenforschung, Darmstadt): Influence of nuclear tracks on the coercivity of single crystalline $(Gd, Bi)_3(Fe, Ga)_5O_{12}$ LPE garnet films.
J. Magn. magn. Mat. **10**, 97-99, 1979 (No. 1). *H*

G. J. P. M. van den Heuvel, G. Jelmorini & G. W. Tichelaar: Messung der Tropfentemperatur im Schweißlichtbogen und ihre Bedeutung in der Praxis.
DVS-Ber. **50**, 165-171, 1978. *E*

R. Hill: A twin line omni-directional aerial configuration.
8th Eur. Microwave Conf., Paris 1978, pp. 307-311. *R*

B. Hoekstra: Spin wave resonance studies of inhomogeneous La,Ga:YIG epitaxial films.
Thesis, Delft 1978. *E*

K. Holford: Capacitance meter.
Wireless World **84**, Oct. 1978, 61-64 (No. 1514). *R*

L. Honds & H. Meyer: Unipolar-Linearmotor mit bewegtem Magneten für verschleißarme Linearantriebe.
Feinwerktechnik & Messtechnik **87**, 152-156, 1979 (No. 4). *A*

P. H. Hooftschrijver (a pseudonym for the user), **F. Meyer, A. H. T. Sanders, H. A. J. Sanders, H. E. M. Mélotte* & H. Bouma*** (* Institute for Perception Research, Eindhoven): Door hoofdbewegingen te bedienen schrijftoestel voor motorisch gehandicapten.
Ned. T. Geneesk. **123**, 460-466, 1979 (No. 12). *E*

J. Hornstra & W. J. Bartels: Determination of the lattice constant of epitaxial layers of III-V compounds.
J. Crystal Growth **44**, 513-517, 1978 (No. 5). *E*

J. B. Hughes, J. B. Coughlin, R. G. Harbott, T. H. J. van den Hurk & B. J. van de Bergh: A 3Gb ECL multiplexer.
1979 IEEE Int. Solid-State Circuits Conf. Dig. tech. Papers, pp. 40-41 + 275. *R*

A. Huijser: Electronic properties of the (110) surface of gallium arsenide and other III-V compounds.
Thesis, Twente 1979. *E*

W. H. de Jeu: The dielectric permittivity of liquid crystals.
Solid State Physics, Suppl. **14**, 109-145, 1978. *E*

W. H. de Jeu: On the viscosity coefficients of nematic MBBA and the validity of the Onsager-Parodi relation.
Physics Letters **69A**, 122-124, 1978 (No. 2). *E*

H. D. Jonker: New method of controlling reproducibility in bubble garnet LPE growth.
IEEE Trans. **MAG-14**, 418-420, 1978 (No. 5). *E*

J. T. C. van Kemenade & R. K. Eijnthoven: Direct determination of barrier voltage in ZnO varistors.
J. appl. Phys. **50**, 938-941, 1979 (No. 2). *E*

G. C. Kenney, D. Y. K. Lou, R. McFarlane, A. Y. Chan, J. S. Nadan, T. R. Kohler, J. G. Wagner & F. Zernike: An optical disk replaces 25 mag tapes.
IEEE Spectrum **16**, Feb. 1979, 33-38. *N*

U. Killat, G. Rabe & D. R. Terrell: Rapid switching of photothermoplastic devices.
Proc. 4th Eur. Electro-Optics Conf., Utrecht 1978 (SPIE Vol. 164), pp. 27-32; 1979. *H*

J. Kittler (University of Oxford) **& P. A. Devijver:** Topics in statistical pattern recognition.
Cambridge Univ. tech. Rep. CAMS 7801, 1978. *B*

R. Knöchel, K. Schünemann (both with T.U. Braunschweig) **& B. Schiek:** Comments on 'The measurement of noise in microwave transmitters'.
IEEE Trans. **MTT-27**, 84-85, 1979 (No. 1). *H*

K. Knorr (Universität Mainz), **W. Gross** (Universität Frankfurt), **J. F. Olijhoek & K. H. J. Buschow:** Crystal fields in $TmCu_5$.
J. Physics F **9**, 645-650, 1979 (No. 4). *E*

H. Koelmans & H. J. Kretschman: Water droplet formation during the life testing of IC's in a humid ambient.
J. Electrochem. Soc. **125**, 1715-1716, 1978 (No. 10). *E*

C. J. Koomen: Information laws for system design.
Proc. Int. Conf. on Cybernetics and Society, Tokyo-Kyoto 1978, Vol. II, pp. 1013-1017. *E*

E. Krätzig: Photorefractive effects and photoconductivity in $LiNbO_3$:Fe.
Ferroelectrics **21**, 635-636, 1978 (No. 1/2/3/4). *H*

E. Krätzig, R. Orlowski, V. Doormann & M. Rosenkranz: Optical information storage in $LiTaO_3$:Fe-crystals.
Proc. 4th Eur. Electro-Optics Conf., Utrecht 1978 (SPIE Vol. 164), pp. 33-37; 1979. *H*

A. E. T. Kuiper, G. E. Thomas & W. J. Schouten: Thin film deposition from beams of ionized atoms and clusters.
J. Crystal Growth **45**, 332-333, 1978. *E*

J. van Landuyt, G. van Tendeloo (both with University of Antwerp, RUCA), **J. J. v.d. Broek, H. Donkersloot & H. Zijlstra:** Defect structure and magnetic properties of MnAl permanent magnet materials.
IEEE Trans. **MAG-14**, 679-681, 1978 (No. 5). *E*

F. H. de Leeuw: An empirical relation for the saturation velocity in bubble domain garnet materials.
IEEE Trans. **MAG-14**, 596-598, 1978 (No. 5). *E*

P. Lesartre, R. Dessert & J. Magarshack: Stable FET local oscillator at 11 GHz with electronic amplitude control.
8th Eur. Microwave Conf., Paris 1978, pp. 264-268. *L*

J. Lohstroh: Static bipolar RAM cell with compact punch-through loads.
1979 IEEE Int. Solid-State Circuits Conf. Dig. tech. Papers, pp. 14-15. *E*

J. Lohstroh: Performance comparison of ISL and I^2L.
1979 IEEE Int. Solid-State Circuits Conf. Dig. tech. Papers, pp. 48-49. *E*

S. R. Longley & P. L. Booth: Frequency tuning of TRAPATT oscillators using ferrimagnetic resonators.
8th Eur. Microwave Conf., Paris 1978, pp. 790-794. *R*

G. Lütteke: Zum Entwurf wirkungsgradoptimierter Mikrowellenoszillatoren mit bipolaren Transistoren.
Thesis, Aachen 1979. *A*

D. Mateika & H. Laudan: Czochralski growth of barium hexaaluminate single crystals.
J. Crystal Growth **46**, 85-90, 1979 (No. 1). *H*

A. Mircea & A. Mitonneau: Strongly anisotropic field ionization of a common deep level in GaAs.
J. Physique Lettres **40**, L 31-33, 1979 (No. 2). *L*

A. E. Morgan & H. W. Werner: Modern methodes for solid surface and thin film analysis.
Physica Scripta **18**, 451-463, 1978 (No. 6). *E*

B. J. Mulder: Fabrication of unbacked ultra-thin films of beryllium and other metals.
Thin Solid Films **55**, 35-40, 1978 (No. 1). *E*

C. Mulder, C. Niessen & R. M. G. Wijnhoven: Layout and test design of synchronous LSI circuits.
1979 IEEE Int. Solid-State Circuits Conf. Dig. tech. Papers, pp. 248-249. *E*

S. G. Nooteboom, J. P. L. Brokx & J. J. de Rooij (Institute for Perception Research, Eindhoven): Contributions of prosody to speech perception.
Studies in the perception of language, ed. W. J. M. Levelt & G. B. Flores d'Arcais, pp. 75-107; Wiley, New York 1978.

T. G. J. van Oirschot, W. J. Leswin, P. J. A. Thijs & W. Nijman: LPE growth of DH laser structures with the double source method.
J. Crystal Growth **45**, 262-266, 1978. *E*

J. A. Pals & J. Dobben: Microwave-induced superconductivity in aluminium strips above the critical temperature.
J. Physique **39**, C6/523-524, 1978 (Colloque C6, Vol. I). *E*

J. A. Pals & J. Dobben: Measurement of characteristic time constants of microwave-enhanced superconductivity in aluminum films.
Phys. Rev. Letters **42**, 270-274, 1979 (No. 4). *E*

J. F. M. Pennings: Adsorption at the interface of a polymer melt and a substrate.
Colloid & Polymer Sci. **256**, 1155-1164, 1978 (No. 12). *E*

G. Perrier, C. Belouet, J. Omaly* & R. Cadoret* (* UER Sciences, Aubière): Vapour growth of HgI_2 in sealed ampoules.
(COSPAR) Space research, Vol. XIX, ed. M. J. Rycroft, pp. 531-534; Pergamon Press, Oxford 1979. *L*

R. J. van de Plassche & R. E. J. van de Grift: A 7b A/D converter.
1979 IEEE Int. Solid-State Circuits Conf. Dig. tech. Papers, pp. 132-133. *E*

J. E. Ralph: Infra-red photoluminescence of CdS:Cu at 10 K.
Solid State Comm. **28**, 377-379, 1978 (No. 5). *R*

W. Rey: A distribution-free estimation of the covariance matrix. Application to the estimation with incomplete data.
Eur. Meeting of Statisticians, Oslo 1978, pp. 166-167. *B*

J. M. Robertson: Liquid phase epitaxy of garnets.
J. Crystal Growth **45**, 233-242, 1978. *E*

D. Rouffet, A. Rabier & B. Kramer: Study of a two carriers solid state straight through repeater at 7.5 GHz.
8th Eur. Microwave Conf., Paris 1978, pp. 620-626. *L*

T. E. Rozzi: Rigorous analysis of the step discontinuity in a planar dielectric waveguide.
IEEE Trans. **MTT-26**, 738-746, 1978 (No. 10). *E*

M. Sarhadi, C. S. Aitchison (both with Chelsea College, London) **& M. J. Underhill:** A fast sampling microwave frequency counter.
8th Eur. Microwave Conf., Paris 1978, pp. 519-523. *R*

H. Schemmann: Zweipolige Einphasen-Synchronmotoren mit dauermagnetischem Läufer. Eigenschaften und Anwendung.
Feinwerktechnik & Messtechnik **87**, 163-169, 1979 (No. 4). *A*

P. C. Scholten: Magnetic birefringence as a tool for determining adsorbed polymer layer thicknesses.
Faraday Disc. Chem. Soc. No. 65, 242-251, 1978. *E*

W. N. Schreiner & R. Jenkins (Philips Electronic Instruments, Inc., Mahwah, N.J.): A non-linear least squares fitting routine for optimizing empirical XRF matrix correction models.
X-ray Spectrom. **8**, 33-41, 1979 (No. 1). *N*

M. Sintzoff: Ensuring correctness by arbitrary post-fixed-points.
Lecture Notes in Computer Science **64**, 484-492, 1978. *B*

E. F. Stikvoort: Increase of gate capacitance in DMOST.
IEEE Trans. **ED-25**, 1388-1394, 1978 (No. 12). *E*

S. Strijbos & P. A. Vermeer (Delft University of Technology): Stress and density distributions in the compaction of powders.
Processing of crystalline ceramics, ed. Hayne Palmour III, R. F. Davis & T. M. Hare, pp. 113-123; Plenum Press, New York 1978. *E*

B. Tell & F. P. J. Kuijpers: Deep-level capacitance spectroscopy of nitrogen-doped VPE GaP.
J. appl. Phys. **49**, 5938-5943, 1978 (No. 12). *E*

J. B. Theeten, D. E. Aspnes* & R. P. H. Chang* (* Bell Laboratories, Murray Hill, N.J.): A new resonant ellipsometric technique for characterizing the interface between GaAs and its plasma-grown oxide.
J. appl. Phys. **49**, 6097-6102, 1978 (No. 12). *L*

G. E. Thomas & E. E. de Kluizenaar: Bombardement-geïnduceerde lichtemissie (BLE); mechanisme en analytische aspecten.
Ned. T. Vacuümtechniek **16**, No. 6, 27-34, 1978. *E*

M. Urner-Wille, T. S. te Velde & P. G. van Engen: Kerr-effect enhancement in amorphous GdFe films.
Phys. Stat. sol. (a) **50**, K 29-31, 1978 (No. 1). *H, E*

M. L. Verheijke & A. W. Witmer: On the calculation of X-ray fluorescence line intensities excited from thin layers on thick substrates.
Spectrochim. Acta **33B**, 817-831, 1978 (No. 10-12). *E*

J. F. Verwey, W. Ruis & I. Sens: Charge transport in oxygen-doped polysilicon layers on Si.
Rev. Phys. appl. **13**, 821-824, 1978 (No. 12). *E*

P. Viktorovitch, G. Kamarinos, P. Even (all with E.N.S.E.R., Grenoble) **& E. Fabre:** Correlation between interface states and MIS silicon solar cell performances.
Phys. Stat. sol. (a) **48**, 137-145, 1978 (No. 1). *L*

Q. H. F. Vrehen & M. F. H. Schuurmans: Direct measurement of the effective initial tipping angle in superfluorescence.
Phys. Rev. Letters **42**, 224-227, 1979 (No. 4). *E*

P. D. White, R. F. Mitchell, R. Stevens, P. Moore* & M. Redwood* (* Queen Mary College, London): Synthesis and design of weighted reflector banks for SAW resonators.
1978 Ultrasonics Symp., Cherry Hill, N.J., pp. 634-638. *R*

F. W. Willmott & I. Mackenzie: Data processing for atomic absorption spectrometry with a microcomputer.
Anal. chim. Acta **103**, 401-408, 1978 (No. 4). *R*

J. P. Woerdman: Effect of a near-resonant intermediate level on Doppler-free versus Doppler-broadened two-photon absorption.
Physics Letters **69A**, 18-20, 1978 (No. 1). *E*

H. E. J. Wulms: Characterisation of I^2L gates.
Microelectronics J. **9**, No. 2, pp. 4-7, 1978. *E*

H. Zijlstra: Trends in permanent magnet material development.
IEEE Trans. **MAG-14**, 661-664, 1978 (No. 5). *E*

Supplement to Philips Technical Review
Beilage der Philips Technischen Rundschau
Bijlage van Philips Technisch Tijdschrift

PHILIPS

December 1981

Recent United States Patents

Abstracts from patents that describe inventions from the following research laboratories, which form part of or cooperate with the Philips group of companies:

Philips Research Laboratories, Eindhoven, The Netherlands	*E*
Philips Research Laboratories, Redhill, Surrey RH1 5HA, England	*R*
Laboratoires d'Electronique et de Physique Appliquée, 3 avenue Descartes, 94450 Limeil-Brévannes, France	*L*
Philips GmbH Forschungslaboratorium Aachen, Weißhausstraße, 51 Aachen, Germany	*A*
Philips GmbH Forschungslaboratorium Hamburg, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany	*H*
Philips Research Laboratory Brussels, 2 avenue Van Becelaere, 1170 Brussels (Boitsfort), Belgium	*B*
Philips Laboratories, N.A.P.C., 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A.	*N*

4 272 574

Optically readable information disc

G. J. M. Lippits *E*
A. J. M. van den Broek
R. Dijkstra

The invention relates to an information disc having a laminated structure which can be read optically. The information disc comprises a transparent substrate which is preferably manufactured from a synthetic resin, for example plexiglass, having thereon a radiation-cured lacquer layer in which the information track is present. The lacquer layer used comprises a radiation cross-linkable protic compound which after curing is aprotic. The lacquer layer preferably comprises a polythiol compound as well as a polyene compound in an equivalent ratio of 1:1.

---

4 272 776

Semiconductor device and method of manufacturing same

B. H. Weijland *E*
W. H. C. G. Verkuijlen

An inset oxide isolated integrated circuit, with multiple levels of inset oxide, polycrystalline regions, and channel stops.

---

4 272 995

Ionization flow meter

M. P. Weistra *E*

An ionization flow meter which can determine with high accuracy the flow of a gas expressed as a gas velocity, volume flow, or mass flow and substantially independently of pressure, temperature and, as the case may be, moisture content. By applying the theory of corona discharges, data relating to the flow-determining factors such as ion mobility and gas density can be derived from the known measured voltage and current values using electronic means. For example, the slope S of the I-V characteristic curve can be determined from the measured values of voltage V and current I.

4 275 091

Method of duplicating plastic information carriers

G. J. M. Lippits *E*
A. J. M. van den Broek
A. J. G. Op het Veld
R. Dijkstra
J. de Jonge

The invention relates to a method of reproducing plastic record carriers, in particular duplicating video records. According to the invention, a metal die is used which is provided with a thin-liquid molding resin of a particular composition which can be polymerized by radiation. A radiation-pervious substrate which is manufactured from synthetic material, for example polymethylmethacrylate, is provided on the molding resin. The molding resin is exposed to light via the substrate after which the cured molding resin together with the substrate connected thereto is removed from the die. The molding resin used in the process comprises low-molecular monomers or oligomers which contain on an average 25-70% by weight of hydrocarbon groups and/or phenyl groups. The molding resin is aprotic and has a functionality as regards unsaturatedness which is between the values 2 and 6. A suitable molding resin contains mono-, tri- or tetra esters of acyclic acid. The molding resin preferably has a swelling capacity with respect to the substrate and for that purpose preferably comprises a vinylmonomer. The metal die used in the method preferably is a quite flat die which is obtained by providing the master disk which is a flat glass plate with information track with a nickel layer, gluing hereon a flat stiffening plate and then removing the master disk. The resulting father disk may be used as a die. Alternatively, further metal copies may be made herefrom which are provided in a simpler manner with a flat stiffening plate. The invention also extends to the molding resin, substrate and die used in the method, as well as to the resulting plastic record carriers.

---

4 276 494

Cathode ray tube with transversely supported electrode and conductive wall coating

J. H. T. van Roosmalen *E*
G. A. H. M. Vrijssen

In a cathode-ray tube, in particular a camera tube, the inner wall of the glass envelope is coated with an electrically conductive material interrupted in the proximity of electrodes extending transversely to

the wall coating and supported by transversely extending supporting surfaces. At the area of each of the interruptions the envelope has a stepwise decrease of the inside diameter in two steps. In the direction of decreasing diameter the first of these steps constitutes the supporting surface for the transverse electrode and the interruption in the wall coating is provided on a wall portion of the second of these steps. The interruptions provided in this manner do not exert any disturbing influence on the electron beam in the tube.

---

4 276 529

Magnet coil arrangement for generating a homogeneous magnetic field for magnetic resonance arrangements

J. Heinzerling *H*
R. Rieckeheer

The invention relates to a magnet coil arrangement for generating a magnetic field which is homogeneous at least in its center, preferably for magnetic resonance spectroscopy. The arrangement consists of four flat ring coils whose coil planes extend perpendicularly to an axis of examination which extends through the coil centers. The coils are symmetrically arranged with respect to a point situated on this axis. By means of a magnetic coil arrangement of this kind, a magnetic field can be generated which extends in the direction of the axis of examination and rotationally-symmetrically with respect thereto and which, in comparison with the magnetic fields generated by means of known coil arrangements, has an improved homogeneity in the direction perpendicular to the axis of examination over a larger range which extends in the direction perpendicular to the axis of examination in the center of the magnetic coil arrangement.

---

4 276 611

Device for the control of data flows

P. G. Jansen *E*
J. L. W. Kessels

A commutation device for the selective control of data transport. At least two data inputs and data outputs, each of the latter having a buffer for storing a data word. A number of possibilities of data transport can be selectively controlled, four for a single connection and two different ones for pair-wise connection. Seven input control lines are provided, two lines for receiving a signal which indicates whether information is present on the associated input line, two lines for indicating the selected output buffer, two erase lines for making a data buffer freely accessible after output of data from the data buffer, and one priority line for granting priority to one of the two input lines if both lines select the same data buffer. There are four output control lines, two lines which indicate that the data present on the input lines have been taken up in the selected output buffer, and two lines which indicate whether an output buffer contains data. The commutation device can effect the data transport itself and can be grouped in specific arrangements to form a buffer in which the data partly determine their own path.

---

4 276 649

Receiver for digital signals in line code

G. C. Groenendaal *E*
E. A. Aagaard

Receiver for a digital line code signal. This receiver comprises a line code decoder and a digital-to-analog converter. To reduce the audibility of bursts this receiver also comprises a line-code violation detector detecting whether the received signal deviates from the line code; as well as a pulse generator. Each time the line-code violation detector detects that the received signal deviates from the line code, the output signal of the pulse generator is applied to the digital-to-analog converter instead of the output signal of the line code decoder.

4 276 650

Method of synchronizing a quadphase receiver and clock synchronization device for carrying out the method

F. de Jager *E*
R. A. van Doorn
J. J. Verboom
M. G. Carasso

The invention relates to a method for the clock synchronization of a receiver for demodulating a quadphase coded data signal and to a clock synchronization device for carrying out the method. In the method according to the invention the first bit is compared (correlated) with the third bit and the second bit with the fourth bit: a high degree of correlation indicates that synchronization has been obtained and a low degree indicates absence of synchronization.

---

4 277 138

Diffraction grating and system for the formation of color components

H. Dammann *H*

A device for spatially separating specific spectral regions, preferably of color components from a wideband spectrum which is actively and/or passively radiated by objects. The spectral regions, or color components, are derived from the diffraction orders of a diffraction grating (phase grating), which is disposed in the pupil of an imaging lens and whose groove profile consists of several steps, which produce path length differences which are integral multiples of a specific wavelength.

---

4 277 542

Resistance material

A. H. Boonstra *E*
C. A. H. A. Mutsaers
F. N. G. R. van der Kruijs

Resistance material consisting of a mixture of metal oxidic compounds, metal oxides, a permanent binder and a temporary binder, the resistance-determining component consisting of barium-rhodate $BaRh_6O_{12}$. This component has a linear positive temperature coefficient of the resistance (TCR) and enables the production of a resistor having a very low TCR by combining the material with a material having a negative TCR. The resistor is obtained by firing this resistance material after it has been applied onto a substrate.

---

4 277 686

Device for determining internal body structures by means of scattered radiation

G. Harding *H*

The invention relates to a device for measuring a scatter coefficient distribution in a plane of a body. The plane is irradiated in different directions by a primary radiation beam along beam paths which are each time situated in parallel in a direction. Scattered radiation which is generated by a primary radiation beam along its path is measured by detectors which are situated on both sides of the plane and which enclose the body as completely as possible. The scatter coefficient distribution is determined by iteration by calculating a scatter value for each beam path from an assumed distribution and by comparing this scatter value with the associated measured scattered radiation. From the difference between calculated and measured values a correction is determined and taken up in the calculated value.

4 277 713

Low-pressure gas discharge lamp and method for making

J. Hasker E
J. C. G. Vervest
C. Peters
L. C. J. Vroomen

Low-pressure discharge lamp having an elongate discharge vessel which contains a thinly distributed filamentary body permeable to the gas discharge, said body comprising a helical support filament which is supported by the inner surface of the discharge vessel and is at least one further filament, supported by the support filament and extending therefrom towards the axis of the discharge vessel.

4 278 888

Apparatus for determining the spatial distribution of the absorption of radiation in a body

W. Wagner H

A computed tomography device wherein a body contour outside an examination area is determined by measurements made with the aid of an auxiliary radiation source (for example light or ultrasound).

4 278 912

Electric discharge tube having a glass-sealed electric leadthrough and method of manufacturing such an electric leadthrough

G. A. H. M. Vrijssen E
J. P. T. Franssen

An electric discharge tube is provided with a hermetically sealed leadthrough which electrically connects electrodes on the inner and outer walls of the envelope. The leadthrough consists of an aperture in the envelope having a conductive layer provided on the wall of the aperture. The aperture is hermetically sealed by means of a plug of thermally devitrified glass which is provided in the form of a suspension of a devitrified glass powder in an organic binder. To manufacture the leadthrough, the envelope of the tube is subjected to temperature treatments in which at a first temperature range the binder is fired from the suspension in an oxygen-containing atmosphere, and at a second temperature range the devitrifiable glass is devitrified in a non-oxidizing atmosphere. A hermetically sealed leadthrough results, without excessive oxidation of the electrodes, while the deformation of the glass envelope at the area of the leadthrough is avoided.

4 279 157

Method of and device for determining the internal structure of a body by means of acoustic beams

H. Schomberg H
M. Tasto

A method and device for determining the internal structure of a body by means of acoustic beams. Transit times and intensities of acoustic beams passing through the body in different spatial directions are measured to establish the refractive index distribution and the acoustic absorption coefficient distribution, respectively at the points of a point matrix associated with the body. The non-rectilinear course of the acoustic beams is taken into account in this respect. This results in reconstructed images of higher quality.

4 279 253

Epilation apparatus

F. Haes E
G. M. P. G. Hermes
C. M. Reijnhout

There is provided an epilation apparatus comprising a drivable member having a hair-gripping wall, and a stationary complementary member having a confronting wall spaced from the hair-gripping wall of the drivable member to provide a hair gap therebetween.

4 280 049

X-ray spectrometer

H. W. Werner E
A. W. Witmer
W. F. Knippenberg

An X-ray spectrometer which is arranged inside an evacuatable housing and which comprises a wavelength dependent X-ray detection system and, for irradiating the specimen to be examined, an electron source with an electron deflection system for generating an electron beam and an X-ray source for generating an X-ray beam. The X-ray source consists of an anticathode on which the electron beam can be directed by the electron deflection system in order to generate the X-ray beam.

4 280 068

Bulk channel charge coupled device having improved input linearity

P. J. Snijder E

In bulk channel charge coupled devices the nonlinearity in the input characteristic caused by varactor effects is removed by moving the potential well in which the charge packets are generated below the input electrode to the surface where the center of electrical charge is substantially independent of the value of the charge. Said shift can be obtained by external means, for example an extra d.c. voltage at the input electrode, or by internal means, for example a thicker oxide below the input electrode.

4 280 089

Automatic incrementing attenuation arrangement

R. J. van de Plassche E
E. C. Dijkmans

Attenuation arrangement comprising a step attenuator arranged in cascade with a controllable voltage divider via first and second voltage terminals, the step attenuator comprising a series arrangement of attenuation elements for dividing a voltage applied across said series arrangement into a plurality of voltage increments, which voltage increments are individually switchable between the two voltage terminals for varying the output voltage of the controllable voltage divider for the voltage range of the relevant voltage element, the direction of the polarity of the voltage between the two voltage terminals changing at a switch-over from one voltage increment to an adjacent voltage increment.

4 280 158

Magnetoresistive reading head

E. de Niet E

A magnetic reading head having a magnetoresistive element which is connected to a reading amplifier. In order to reduce the modulation noise (Barkhausen effect) when making the relationship between the resistance variation and the strength of the signal field linear in a negative feedback loop of the reading amplifier an electric turn is present which turn is positioned relative to the magnetoresistive element in such manner that a negative feedback field (Ht) can be generated with it which causes a magnetic flux in the element which is directed oppositely to the magnetic flux caused in the element by a magnetic field (H_y) to be detected.

4 280 858

Method of manufacturing a semiconductor device by retarding the diffusion of zinc or cadmium into a device region

C. J. M. van Opdorp E
H. Veenvliet

A semiconductor device and a method for manufacturing the semiconductor device are disclosed for forming an abrupt and accurately positioned p-n junction between a substrate and a substrate-adjoining region. This is achieved in accordance with the present

invention by diffusing zinc or cadmium from a surface of the substrate-adjoining region to the substrate, and abruptly limiting or retarding the diffusion of the zinc or cadmium into the substrate near a junction between the substrate and the region. This is accomplished in accordance with the present invention by selecting the net donor concentration in the substrate near the junction to be higher than the concentration of zinc or cadmium at the surface of the substrate-adjoining region.

---

4 281 396

Magnetic strip domain memory system

J. Roos E

A magnetic memory device in which information is stored in the form of strip domains in a layer of magnetic material supported by a layer of ferromagnetic material. The ferromagnetic material contains a pattern of alternately magnetized strips for sustaining a magnetic field periodically varying in a first coordinate direction and directed transverse to the domain layer. The device also includes a generator for receiving and converting data into configurations of the strip domains in the plate.

---

4 283 226

Method of preparing titanium iron-containing material for hydrogen storage

H. H. van Mal E
H. A. van Esveld
J. S. van Wieringen
K. H. J. Buschow

A material for storing hydrogen consisting of a titanium-iron alloy having 5-30 at.% of one or more metals of the group chromium, zirconium, manganese and vanadium.

---

4 283 689

Microwave oscillator circuit with improved efficiency

H. Tjassens E

A microwave oscillator circuit, suitable for use as the local oscillator in beam transmitters, radar systems and satellite TV receivers, comprises an active element (IMPATT, Gunn diode) at one end of a coaxial transmission line which is terminated at its other end by a matched load. At a suitable distance from the diode a high Q transmission cavity resonator is coupled to the transmission line via a first coupling hole. A drawback of such a circuit is that a portion of the oscillator power at the required oscillator frequency f_o is dissipated in the terminal impedance Z_o. This is obviated by coupling the transmission resonant cavity to the transmission line via a second coupling hole. The distance between the first and the second coupling hole is $\frac{1}{4}\lambda$. As a result the terminal impedance, which has been transformed very frequency-selective to a very high value in situ of the second coupling hole, is transformed to a very low value at the first coupling hole and very little power is dissipated in this low impedance, so that a considerable improvement of the circuit efficiency has been achieved.

---

4 283 837

Semiconductor device and method of manufacturing same

A. Slob E

A semiconductor device includes a silicon substrate having an insulating layer with a window. A silicon layer is deposited on the insulating layer and on the silicon substrate surface in the window. This silicon layer has n-type and p-type conductive layer parts which adjoin each other within the window and which each serve as both a connection conductor and an electrode of an active zone of the device. Semiconductor devices in accordance with the invention feature very small surface areas, and are thus particularly suitable for high frequency operation.

4 284 069

Wall element comprising a solar collector which is disposed between two transparent panes

H. Hörster A
W. Hermann
K. Klinkenberg

A wall element, comprising a solar collector which is arranged between two panes and which comprises a number of rotatable absorber plates. One side of the absorber plates is provided with a non-selective black coating and the other side is provided with a selective, heat-reflective layer. The absorber plates are accommodated in evacuated, transparent tubes.

---

4 286 156

Device for determining the spatial absorption distribution in a plane of examination

W. Wagner H

The device in accordance with the invention comprises detectors, a first part of which is not struck by radiation during measurement of the useful signal, whilst a second part which is struck directly by the radiation during the measurement of the useful radiation, is shielded during a next measurement. During the last measurement, the first detector part is struck by scattered radiation. From the two signals thus formed, a signal free from scattered radiation can be obtained by subtraction.

---

4 286 177

Integrated injection logic circuits

C. M. Hart E
A. Slob

An "Integrated Injection Logic" integrated circuit in which bias currents are supplied by means of a current injector. The current injector is a multi-layer structure in which current is supplied by means of injection and collection of charge carriers via rectifying junctions, to predetermined zones of the circuit to be biased. Such zones are preferably biased by charge carriers which are collected by such zones from one of the layers of the current injector. The circuit also preferably includes a region for reducing carrier injection from a predetermined zone.

---

4 286 266

Display device

M. de Zwart E
J. L. A. M. Heldens

In a liquid crystal display device which upon controlling with direct voltage shows a memory effect that can be erased by alternating voltage, the written area can expand beyond the edge of the electrode so that edge zones can no longer be erased readily. This disadvantage is avoided by covering the edges of the electrodes with an insulating layer.

---

4 286 318

Control loop

K. A. Immink E
A. Hoogendoorn

A control loop provided with a control unit for realizing a transfer characteristic having a number of peaks at a fundamental frequency and harmonics thereof. The control unit comprises a memory device for digitally storing a number of samples of the error signal appearing in the control loop during a cycle period equal to the period corresponding to the fundamental frequency. Furthermore, there are provided means for comparing the sample stored in the memory device with the value of the error signal one cycle period later and, depending on this comparison, correcting the memory content of the relevant memory location. The variation of the error signal stored in the memory device is furthermore cyclically employed as a control signal for the control loop.