PHILIPS TECHNICAL REVIEW



VOLUME 36
1976

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: Drs J. W. Miltenburg, Editor-in-chief
Dr J. W. Broer
Ir D. van Dalen
Dr E. Fischmann
Drs K. J. Groenewolt
Ir J. A. Klaassen
Dr J. Ubbink
M. B. Verrijp, assistant editor

English edition: D. A. E. Roberts, B.Sc., M.Inst.P., A.I.L. (Redhill, Surrey)

German edition: Dipl.-Phys. R. Dockhorn (Hamburg)

© N.V. Philips' Gloeilampenfabrieken, Eindhoven, The Netherlands, 1977.
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 36, 1976

	Page
Acoustic surface-wave bandpass filters	29
Actuator, fast, for anti-lock braking system	74
Alkali-vapour dispenser, easily controlled	16
Alloys, heat of formation of	217
Alumina, sintered, light transmission of	47
Anti-lock braking system, fast actuator for	74
Avalanche photodiode as detector for optical communication via glass-fibre cables	205
Braking system, anti-lock, fast actuator for	74
Bubble memory, single-mask, with rotating-field control	149
CAD/CAM system	162
Carbon foam	93
Circulators, broadband, for VHF and UHF	255
Computer-aided design	162
Computer simulation in designing the control system for the Amsterdam Water-Supply Authority	273
Cooling, continuous, in the millikelvin range	104
Data transmission	349
Deflection, electro-optic, of a laser beam	117
Dilution refrigerator	104
Dispenser for sodium, easily controlled	16
Dispersion measurements on optical fibres, test bench for	211
Domains in ferroelectrics and ferromagnetics, SEM observations of	18
Electro-optic deflection of a laser beam	117
Electrostatic printing	57
Ergonomic lathe	160
Ferroelectrics) SEM observations of domains	
Ferromagnetics) in	18
Filters, acoustic surface-wave	29
Glass fibres, see **Optical communication via glass-fibre cables**	
Heat of formation of alloys	217
High-pressure apparatus for solid-state research	245
Intermetallic compounds; background and results of twenty years of research	136
Langmuir trough, automated, for building monomolecular layers	44
Laser beam, electro-optic deflection of	117
Lasers, semiconductor, for optical communication via glass-fibre cables	190
Lathe, ergonomic	160
Light modulation and injection in optical communication via glass-fibre cables with semiconductor lasers	201
Light transmission of sintered alumina	47
Loudspeaker cones, vibration patterns and radiation behaviour of	1
Milk, experimental monitor for	284
Modulation:	
of light in optical communication	201
general survey (special issue, see contents on page 362)	305
Monomolecular layers, building of, with automated Langmuir trough	44
Optical communication via glass-fibre cables:	
systems	178
manufacture of the glass fibres	182
semiconductor lasers	190
light modulation and injection	201
avalanche photodiode as detector	205
test bench for characterizing the fibres by pulse dispersion	211
Pesticides in milk, organochlorine, experimental monitor for	284
Printing, electrostatic	57
Quantization and coding of analog signals	337
Scanning-electron-microscope (SEM) observations of domains in ferroelectrics and ferromagnetics	18
Single-mask bubble memory with rotating-field control	149
Solid-state research at very high pressures, apparatus for	245
Surface-water purification plant for the Amsterdam Water-Supply Authority, control of	273
Surface waves, acoustic, application in filters	29
Suspension technology	264
Telephone and video-telephone signals, switching of	291
Tungsten transport by the water cycle	133
Video telephone:	
experimental system	85
digital circuits	233
switching of signals	291
Visual conspicuity, measurement of	71
Water cycle, tungsten transport by	133
Water supply, Amsterdam: control of surface-water purification plant	273

Author index, Volume 36, 1976

Page

Aagaard, E. A., P. M. van den Avoort and F. W. de Vrijer
An experimental video-telephone system 85
Acket, G. A., J. J. Daniele, W. Nijman, R. P. Tijburg and P. J. de Waard
Semiconductor lasers for optical communication . . . 190
Ass, H. M. J. M. van, P. Geittner, R. G. Gossink, D. Küppers and P. J. W. Severin
The manufacture of glass fibres for optical communication 182
Avoort, P. M. van den, see Aagaard, E. A.
Blume, P.
Computer-aided design 162
Bollen, L. J. M., J. J. Goedbloed and E. T. J. M. Smeets
The avalanche photodiode 205
Boudewijns, H. P. J., E. C. Dijkmans, P. W. Millenaar, N. A. M. Verhoeckx and C. H. J. Vos
Digital circuits in the video telephone 233
Brouha, M. and A. G. Rijnbeek
Apparatus for solid-state research at very high pressures 245
Coenders, J. W.
Switching telephone and video-telephone signals . . . 291
Daniele, J. J., see Acket, G. A.
Davis, G. L.
Transport of tungsten by the water cycle 133
Dijkmans, E. C., see Boudewijns, H. P. J.
Dolphin, R. J., L. P. J. Hoogeveen and F. W. Willmott
An experimental system for the automatic determination of organochlorine residues in milk 284
Druyvesteyn, W. F., F. A. Kuijpers, A. G. H. Verhulst and C. H. M. Witmer
Single-mask bubble memory with rotating-field control 149
Engel, F. L.
The measurement of visual conspicuity 71
Engelsen, D. den, J. H. Th. Hengst and E. P. Honig
An automated Langmuir trough for building monomolecular layers 44
Frankort, F. J. M.
Vibration patterns and radiation behaviour of loudspeaker cones 1
Fransen, J. J. B. and J. H. N. van Vucht
An easily controlled alkali-vapour dispenser 16
Frens, G., H. F. Huisman, J. K. Vondeling and K. M. van der Waarde
Suspension technology 264
Geittner, P., see Ass, H. M. J. M. van
Goedbloed, J. J., see Bollen, L. J. M.
Gossink, R. G., see Ass, H. M. J. M. van
Hazan, J.-P. and L. Jacomme
Characterizing optical fibres; a test bench for pulse dispersion 211
Hengst, J. H. Th., see Engelsen, D. den
Honig, E. P., see Engelsen, D. den
Hoogeveen, L. P. J., see Dolphin, R. J.
Huisman, H. F., see Frens, G.
Jacomme, L., see Hazan, J.-P.
Khoe, G. D. and L. J. Meuleman
Light modulation and injection in optical-fibre transmission systems with semiconductor lasers 201

Page

Klinck, M.
Control of the surface-water purification plant for the Amsterdam Water-Supply Authority 273
Knippenberg, W. F. and B. Lersmacher
Carbon foam 93
Kuijpers, F. A., see Druyvesteyn, W. F.
Küppers, D., see Ass, H. M. J. M. van
Lersmacher, B., see Knippenberg, W. F.
Meuleman, L. J., see Khoe, G. D.
Michel, C.
Observations of domains in ferroelectrics and ferromagnetics with a scanning electron microscope 18
Miedema, A. R.
The heat of formation of alloys 217
Millenaar, P. W., see Boudewijns, H. P. J.
Mouthaan, K.
Optical communication systems with glass-fibre cables . 178
Nijman, W., see Acket, G. A.
Parker, D. W., R. G. Pratt, F. W. Smith and R. Stevens
Acoustic surface-wave bandpass filters 29
Peelen, J. G. J.
Light transmission of sintered alumina 47
Pratt, R. G., see Parker, D. W.
Rijnbeek, A. G., see Brouha, M.
Rothgordt, U.
Electrostatic printing 57
Schiefer, G.
Broadband circulators for VHF and UHF 255
Schmidt, U. J.
Electro-optic deflection of a laser beam 117
Severin, P. J. W., see Ass, H. M. J. M. van
Skoyles, D. R.
A fast actuator for an anti-lock braking system 74
Smeets, E. T. J. M., see Bollen, L. J. M.
Smith, F. W., see Parker, D. W.
Staas, F. A.
Continuous cooling in the millikelvin range 104
Stevens, R., see Parker, D. W.
Tijburg, R. P., see Acket, G. A.
Verhoeckx, N. A. M., see Boudewijns, H. P. J.
Verhulst, A. G. H., see Druyvesteyn, W. F.
Vondeling, J. K., see Frens, G.
Vos, C. H. J., see Boudewijns, H. P. J.
Vrijer, F. W. de
Modulation 305
I. Modulation of a sinusoidal carrier 309
II. Modulation of pulse trains 329
III. Quantization and coding of analog signals 337
IV. Transmission of digital signals 343
V. Modulation in telecommunication 353
Vrijer, F. W. de, see Aagaard, E. A.
Vucht, J. H. N. van
Intermetallic compounds; background and results of twenty years of research 136
Vucht, J. H. N. van, see Fransen, J. J. B.
Waard, P. J. de, see Acket, G. A.
Waarde, K. M. van der, see Frens, G.
Willmott, F. W., see Dolphin, R. J.
Witmer, C. H. M., see Druyvesteyn, W. F.

PHILIPS TECHNICAL REVIEW

VOLUME 36, 1976, No. 1

Vibration patterns and radiation behaviour of loudspeaker cones

F. J. M. Frankort

Many treatments of loudspeakers begin with the comment that the loudspeaker is the weakest link in the audio reproduction chain. The complex vibrational behaviour of the conical loudspeaker diaphragm, which noticeably affects the frequency response of the loudspeaker, has been a frequent challenge to mathematical analysis. Only recently however, with the availability of the computer to carry out extensive numerical computations, has it been possible to obtain any detailed picture of the behaviour of the loudspeaker cone. The picture gives a satisfactory explanation of the irregularities in the behaviour of the cone as a radiator, which had long been known from acoustic measurements. At the same time it has now become possible to indicate the dimensions and material properties that will produce the desired frequency response.

Introduction

The behaviour of the conical diaphragm, a shape that was intuitively chosen for loudspeakers right from the beginning, is rather complex. The audible sound spectrum contains widely different frequencies (about 16 Hz to 20 000 Hz); when alternating currents of these frequencies are fed to a single loudspeaker, the diaphragm will be caused to vibrate in different modes of lower and higher order. It is only at low frequencies that the cone vibrates as a rigid body. It is not stiff enough to withstand the inertial forces that occur at higher frequencies; it starts to vibrate in parts and the cone is said to 'break up'. The higher-order modes of vibration that now appear enable the loudspeaker to fulfil its function at higher frequencies and set the air in motion.

The amount of the air displacement depends to a great extent on whether the cone is caused to vibrate at its resonant frequencies. Hence the marked variation of the sound radiation as a function of frequency. As can be seen in *fig. 1*, measurement of the pressure response with a microphone at some distance from the loudspeaker produces a rather irregular curve.

In the ideal case the sound radiation would have the same amplitude at all frequencies, and the frequency response would be linear. Loudspeaker cones that approximate to this requirement have hitherto been designed mainly on empirical lines. An efficient design procedure requires detailed knowledge of the radiation behaviour and its effect on the properties of the cone material. The necessary detailed information can be obtained by setting up the differential equations that

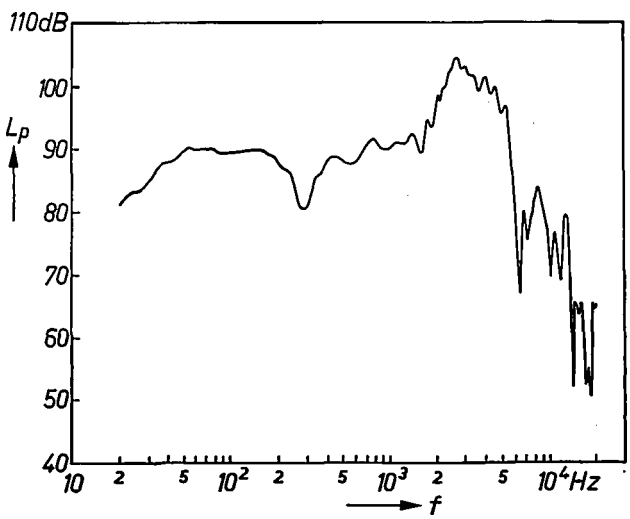


Fig. 1. Example of the frequency characteristic of a loudspeaker. The sound-pressure level L_p is measured as a function of the frequency f at a distance of 10 metres. The loudspeaker was mounted in a baffle.

Dr Ir F. J. M. Frankort, now teaching at the Hogere Technische School, Heerlen, was formerly with Philips Research Laboratories, Eindhoven.

describe the cone vibrations and then, in the absence of an analytical solution, solving them numerically for a large number of frequencies. This is now possible with the computer.

In this way frequency and directivity characteristics can be calculated. We have made such calculations and found that the general shape of the curves agrees well with the measured curves [1]. This agreement largely depends on the correct prediction of the various natural frequencies of the cone. Once these are known, together with various other characteristic frequencies, the shape of the frequency response can be broadly predicted. Conversely, in designing a loudspeaker cone, it is possible to derive the locations of the characteristic frequencies from the shape of the desired characteristic.

To obtain a good understanding of the numerical results it is necessary to take into account the behaviour of the longitudinal and transverse waves on an elastic cone. Before presenting the results of the calculations, we shall therefore give a general picture of this behaviour when certain simplifications are introduced. The complexities in the vibrational behaviour of an elastic conical diaphragm are of course not encountered in the hypothetical case of a diaphragm in which every point describes the same movement, so that it moves to and fro like a rigid piston. This hypothetical case will serve to introduce some important concepts and define some characteristic frequencies.

Rigid cone

Fig. 2 shows a cross-section of a typical loudspeaker construction. The conical diaphragm *D* is flexibly mounted by means of an outer suspension or rim *OS* and an inner suspension *IS*. This method of suspension only allows an axial motion. The drive force is supplied by the voice coil *VC*, which moves in the air gap of a permanent magnet *M*. The mass of the cone and voice coil and the stiffness of the suspension form the elements of a spring-mass system, whose resonant frequency is f_0.

Above this frequency the alternating drive force mainly serves to overcome the inertia of cone and voice coil. If the force is the same at all frequencies the amplitude of the acceleration will also be the same at all frequencies; the velocity decreases with increasing frequency.

What consequences does this have for the sound radiation? To answer this question we calculate the sound pressure that the movements of the diaphragm produce at a point some distance away from the loudspeaker. We treat the vibrating diaphragm here as a collection of point sources uniformly distributed over the surface, and we add together the contributions

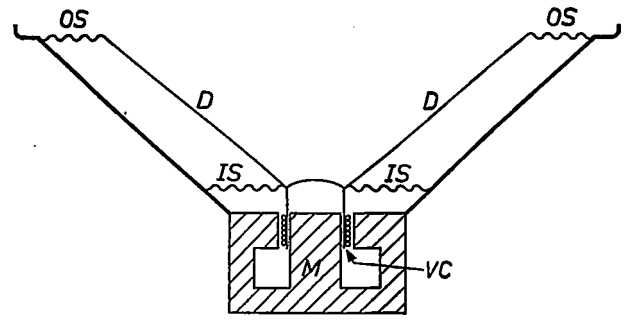


Fig. 2. Cross-section of a loudspeaker. *D* diaphragm. *OS* outer suspension. *IS* inner suspension. *VC* voice coil. *M* permanent magnet.

from all these point sources. For simplicity we first consider the diaphragm as a flat piston and assume that it vibrates in an infinite baffle. Assuming that the amplitude of the piston velocity is fixed, we then find that the sound pressure at the point of observation increases linearly with the frequency. This increase exactly compensates for the velocity decrease due to inertia, and the result for a fixed drive force is thus a fixed frequency-independent sound pressure.

This is valid above f_0, but only for low frequencies. The pathlength to the point of observation is not the same for all the individual point sources on the piston, and therefore their contributions to the total sound pressure do not arrive exactly in phase. In the case of very long waves the differences in pathlength is not significant, but at shorter wavelengths, i.e. higher frequencies, it leads to phase differences that cannot be neglected. These are greater for radiation to the sides, so that the piston does not radiate the same power in all directions; at higher frequencies the piston exhibits a directional effect. The forward radiation is the strongest, and therefore sound-pressure measurements are nearly always made with the microphone on the axis of symmetry of the loudspeaker.

When a frequency characteristic is recorded with a microphone on the axis of the piston, nothing is noticed of the directivity at higher frequencies (see the frequency characteristic in *fig. 3a*, dashed curve). However, when the total sound power radiated in all directions is measured (e.g. by using several microphones) a decrease is observed when the directivity starts to appear (fig. 3*b*, dashed curve). The transition is gradual, but for practical reasons we define a transition frequency f_t. When the horizontal and sloping parts of the dashed curve in fig. 3*b* are extended, they intersect at this frequency. The sound wavelength at this frequency is found to be approximately equal to the circumference of the piston.

We shall now go a step further and assume that the rigid piston has the shape of a loudspeaker cone. The

calculation of the sound radiation now becomes more complicated. At higher frequencies, where the depth of the cone is no longer negligible compared with the wavelength, or may even be greater than the wavelength, the radiation deviates from that of the flat piston. The radiation from different parts of the cone then arrives at the point of observation with appreciably different phases, even when the point is on the axis of the loudspeaker. This results in a lower sound pressure at this point (see fig. 3*a*, solid curve). For the beginning of this decrease a cut-off frequency f_c is defined, at which the sound wavelength is about three times the cone depth. Cones with a conventional apex angle have an f_c greater than f_t.

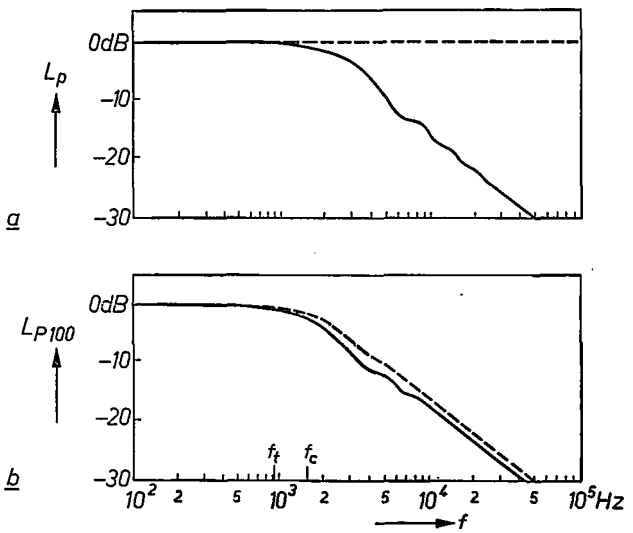


Fig. 3. *a*) Frequency characteristic of the sound-pressure level L_p with a rigid piston (*dashed curve*) and with a rigid cone (*solid curve*). *b*) Frequency characteristic of the total sound power radiated inside a conical region of apex angle 100°. f_t frequency above which a rigid piston gives a directional effect. f_c frequency above which pathlength differences due to the depth of the cone reduce the sound radiation.

Flexible cone

In reality a loudspeaker cone is by no means a rigid body. Above certain frequencies both transverse and longitudinal waves appear in the conical shell. These waves are coupled and together determine the vibration pattern, which considerably affects the air displacement. If we compare the measured frequency characteristic in fig. 1 with the calculated solid curve in fig. 3*a*, we see from the actual characteristic that the loudspeaker will function up to much higher frequencies.

In the discussion that now follows of the coupling between transverse and longitudinal waves we shall encounter a rather interesting resonant mode, which does not occur in a flat plate, and has an important bearing on the behaviour of the cone as a radiator.

Two types of wave

In a flat plate the transverse and longitudinal waves do not affect one another. Both types of wave propagate faster the stiffer the medium. Since the plate is much stiffer for longitudinal compression and expansion than for bending, the longitudinal waves are much longer than the bending waves.

In a cone the situation is more complicated. In general the two wave motions cannot exist independently. A displacement normal to the cone surface leads to a displacement along the surface of the cone, and vice versa. This may be illustrated for a conical ring, on the inner edge of which a uniformly distributed longitudinal force F_l is exerted (*fig. 4a*), giving a longitudinal displacement u. An increase in the diameter of the ring is therefore implied; as a result of the extension a tension appears in the ring, directed at all points along the tangent to the circumference. Considering a segment of the ring (fig. 4*b*) we see that the tensile forces at the two ends of the segment result in a force F_c directed towards the centre of the ring. This force lies in the plane of the ring (fig. 4*c*) and can be resolved into

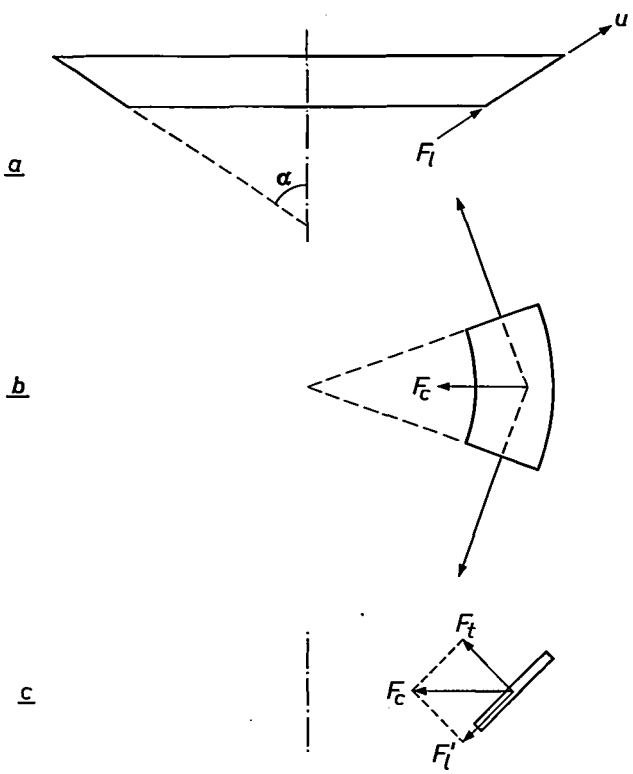


Fig. 4. Longitudinal displacements in a cone set up transverse forces, and vice versa. *a*) Side view of a conical ring (semi-apex angle α), on which a distributed force F_l acts in the longitudinal direction. u longitudinal displacement. *b*) The longitudinal displacement sets up a tension in the ring. The tensile forces at the ends of a part of the ring result in a force F_c directed towards the centre. *c*) Cross-section. F_c has a transverse component F_t and a longitudinal component F_l'.

[1] An extensive treatment is given in F. J. M. Frankort, Vibration and sound radiation of loudspeaker cones, Thesis, Delft 1975, also published as Philips Res. Repts. Suppl. 1975, No. 2.

a transverse component F_t and a longitudinal component F_l'. In the equilibrium position F_l' is equal and opposite to F_l. The applied longitudinal force F_l thus gives rise to a transverse force F_t, whose magnitude depends on the apex angle of the cone.

Both types of wave propagate from the apex of the cone to the edge and back as well as in circular paths parallel to the circumference of the cone. The outward-travelling waves are reflected at the edge and also at the point where the drive is applied. At certain frequencies standing waves occur (*fig. 5a*), and the nodal lines then appear as concentric circles. These standing waves occur only at higher frequencies and greatly affect the sound radiation.

In the case of wave propagation in circular paths, standing waves occur at frequencies where a circular path is exactly two or more wavelengths long (fig. 5*b*). Nodal lines then appear along generatrices of the cone. Since the bending stiffness of the cone is relatively small for circulating waves, the propagation velocity is low and these standing waves appear even at low frequencies, e.g. at about 100 Hz for a typical 20-cm (8-inch) loudspeaker. They have very little effect on the sound radiation since the parts of the cone moving in antiphase are so close together that at these low frequencies their effects are practically cancelled out, an effect known as 'acoustical short-circuiting'. This mode of vibration will not therefore be considered here.

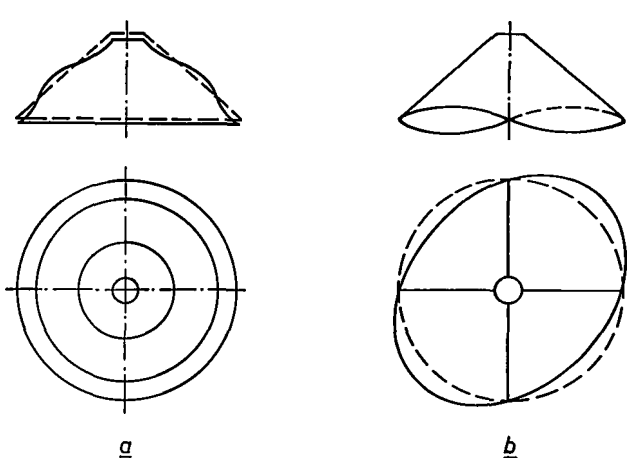


Fig. 5. Cone with standing-wave pattern of *a*) outgoing and reflected waves, *b*) circulating waves. The vibrational modes in (*b*) radiate little sound.

An interesting resonance effect

If the loudspeaker diaphragm is a truncated cone, it has both an inner and an outer edge; the drive force is applied to the inner edge. When the waves propagated from the inner edge return to it in phase after reflection from the outer edge they increase the amplitude of the displacement there; this is referred to as a resonance. There will also be frequencies, however, for which the waves return to the inner edge in antiphase and oppose the displacement; here we have an antiresonance. This holds both for longitudinal and for bending waves and it applies not only for a cone but also for a flat ring.

In the case of a cone the coupling between longitudinal waves and bending waves gives rise to an antiresonance that is not encountered in a flat ring. The frequency at which this antiresonance occurs is a characteristic frequency of the cone: bending waves only occur above this frequency.

To explain this special mode of vibration we first consider a narrow conical ring to which, as in fig. 4, a longitudinal force F_l is applied; see *fig. 6*. We now assume, however, that F_l is a sinusoidally alternating force, which implies that inertia will come into the picture.

To begin with we can consider that the ring has a characteristic mode of vibration in which it alternately contracts and expands while retaining its shape; the ring vibrates in a plane perpendicular to the axis and its centre of gravity remains at rest (fig. 6*a*). The elasticity of the ring provides the spring component of the spring-mass system. The mode occurs at the ring resonant frequency f_{rr}. If the drive force F_l alternates at this frequency, then the longitudinal displacements u and the transverse displacements w would both become infinitely large in the absence of damping, and so would the amplitude of the velocity in both directions. The mechanical impedance, defined as force divided by velocity, would then be zero at this frequency.

This mode of vibration is not specific to a conical ring; it also occurs in rings of other cross-sections. However, in the case of the antiresonance mentioned above, which is confined to a conical ring and occurs only at a frequency $f_{ra} < f_{rr}$, a vibration is found at which, in spite of the longitudinal drive, the longitudinal amplitude is small (in the theoretical case with no damping it is in fact zero; fig. 6*b*). We call the frequency f_{ra} the ring antiresonant frequency. At this frequency the force F_l encounters a high mechanical impedance (*infinitely* high with no damping). Nevertheless, this antiresonance is essentially longitudinal, since the circumference of the ring becomes alternately larger and smaller. This longitudinal antiresonance can also be considered as a transverse *resonance* of the ring mass and the transverse component of the spring formed by the elasticity of the ring. In this form of motion, however, the centre of gravity of the ring is *not* at rest. In other words, instead of a free vibrational mode of the ring we have a forced vibration, which can only occur when the drive is longitudinal as described above. The axial component of the drive produces the movement of the centre of gravity, in accordance with Newton's laws of motion.

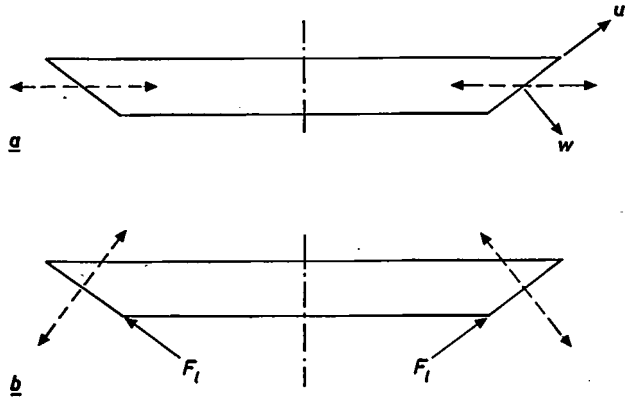


Fig. 6. Mode of vibration of a conical ring for *a*) resonance, *b*) antiresonance. u longitudinal displacement, w transverse displacement. At the antiresonance the longitudinal displacement is zero even though the drive is provided by a longitudinal force F_1.

The resonance effects described related to a ring. We shall now, by way of approximation, consider the loudspeaker cone as a set of coupled conical rings. Each ring can be considered as a concentrated mass and a spring, so that the complete cone can be much more simply represented by a system of masses and springs (*fig. 7*). Each ring can be excited at its antiresonant frequency f_{ra}. The transverse amplitude then becomes large.

The frequency f_{ra} is lowest for the outer ring, since this has the largest mass and the weakest spring. As the frequency increases, the resonance effect gradually travels inwards, and eventually arrives at the inner edge. The ring in resonance marks a transition circle. On the stiffer part of the cone, inside this circle, the waves that occur are mainly longitudinal; on the part outside it, which is less rigid, they are mainly bending waves. On the transition circle itself there is an exchange of energy between longitudinal and bending waves.

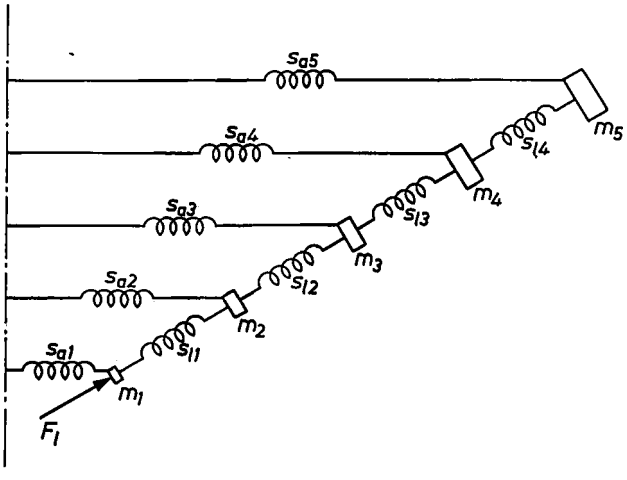


Fig. 7. Representation of a cone as a system of masses m and springs with stiffnesses s_l in the longitudinal direction and s_a in the azimuthal direction. The values of m increase the further they are from the centre, and the values of s_a decrease.

The frequency at which the resonance at the outer edge begins is called f_{to}. The resonance reaches the inner edge at the frequency f_{ti}; the entire cone is then covered with bending waves.

In the vibrational behaviour of a cone the longitudinal displacement u_i at the inner edge is of considerable importance. If the cone consisted of only one ring, this displacement would be zero at the ring antiresonant frequency f_{ra}. An actual cone can be approximated by a ring of radius equal to the outer radius R_o of the cone, which is connected by a relatively stiff part to the inner edge of the cone. The stiff part transmits the force to the outer ring but adds mass to the system. For the cone as a whole, therefore, the frequency f_{ra} at which u_i is zero is lower than the antiresonant frequency f_{to} of the outer ring itself. The difference, however, is not very great, and instead of f_{ra} it may often be more convenient to use f_{to}, whose value is much easier to calculate. The value of f_{ra} can only be calculated numerically from the equations of motion of the complete cone.

Calculation of the cone vibrations

The representation given in fig. 7 is of course a considerable simplification. In reality, there are not only tensile stresses acting on an element of the conical shell, but bending moments as well. There is also damping, caused by the radiation of acoustical energy and also by internal losses in the material and suspension of the cone. If we wish to take all these factors into account, the simple models used so far are inadequate. We then have to resort to a complete mathematical treatment. A description that takes all the forces and moments into account, but not the damping, has eight simultaneous first-order differential equations with eight unknowns.

Disregarding the vibrational modes with diametric nodal lines (fig. 5*b*) since they make no contribution to the sound radiation, we can reduce this set to six simultaneous differential equations. The six unknowns are all expressed as a function of the coordinate x along a generatrix of the cone (at the apex of the cone x is zero, at the inner edge a, at the outer edge b). For the solution of the six equations three boundary conditions are imposed at each edge of the cone. The equations are solved by direct numerical integration [2], carried out for a large number of different frequencies. If the damp-

[2] First applied to these problems by J. E. Goldberg, J. L. Bogdanoff and L. Marcus (On the calculation of the axisymmetric modes and frequencies of conical shells, J. Acoust. Soc. Amer. 32, 738-742, 1960) and refined by A. Kalnins (Analysis of shells of revolution subjected to symmetrical and nonsymmetrical loads, Trans. ASME E (J. appl. Mech.) 31, 467-476, 1964; Free vibration of rotationally symmetric shells, J. Acoust. Soc. Amer. 36, 1355-1365, 1964).

ing is included, all the fundamental variables become complex numbers, in which case we have twelve differential equations and twelve boundary conditions. In some of the results we shall later present the damping is taken into account.

The boundary conditions follow directly from the loudspeaker construction. The connection between the inner edge of the cone and the voice coil consists of a stiff rim strengthened by adhesive, and the inner suspension ensures that only axial motion is possible. We therefore assume in the calculations that the inner edge of the cone is clamped to an infinitely stiff ring of zero mass and that the movements of the ring are purely axial. The purely axial movement implies a fixed ratio between transverse and longitudinal amplitudes at the inner edge; this is one of the boundary conditions. Another boundary condition relates to the clamping of the inner edge ($\partial w/\partial x = 0$); the third boundary condition is that the forces at the inner edge are in equilibrium. The boundary conditions for the outer edge are that it has freedom of movement, implying that all forces and moments there are zero.

It can be shown that the mechanical impedance to the axial drive force is given by:

$$Z_a = Z_t \sin^2\alpha + Z_l \cos^2\alpha.$$

In this expression Z_t is the transverse and Z_l the longitudinal impedance that would be present at the inner edge of the cone in the absence of the stiff ring. The transverse impedance, which is connected with the bending stiffness of the cone, is generally much smaller than the longitudinal impedance, which is determined by expansion and compression in the plane of the conical shell. It is primarily Z_l, therefore, that determines the amplitude of the displacements of the inner edge.

A diagram of the situation is shown in *fig. 8*, where the impedances Z_t and Z_l are represented by springs with the stiffnesses s_t and s_l. It can be seen that if the stiffness s_t is much smaller than s_l, the axial motion depends mainly on s_l.

The outer suspension is not included in the calculations. The damping it normally introduces is accounted for by using a higher internal damping in the calculations than the actual value for the cone material.

The vibration patterns in different frequency regions

We have calculated the transverse and longitudinal displacements of a cone with characteristics like those encountered in practice. We shall refer to this as cone 50.1; the number 50 indicates the semi-apex angle in degrees (the geometry and material parameters are given in *Table I*). For simplicity the internal damping is assumed to be zero ($D = 0$).

From the transverse amplitude $w(x)$ and the longitudinal amplitude $u(x)$ we have plotted the magnitude and direction of the resultant motion for a number of points on the cone. The vibration patterns of a cone cross-section obtained in this way are shown in *fig. 9a-i*. Three frequency regions can be distinguished. The first (region I) contains the frequencies up to the appearance

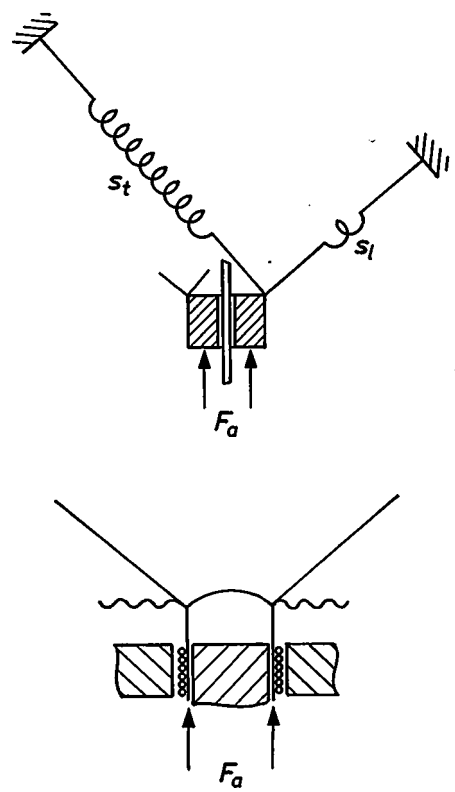


Fig. 8. Illustrating the mechanical impedance to an axial force F_a on the inner edge of a cone. The impedance is represented by the simultaneous action of two springs. s_t impedance to transverse displacements. s_l impedance to longitudinal displacements. The inner edge is capable of axial motion only. In practice the longitudinal impedance predominates and mainly determines the displacements.

Table I. Dimensions and material constants of the loudspeaker cones mentioned in the article. The letter e after a number indicates an experimental cone; the others are calculated examples. α semi-apex angle. R_i inner-edge radius. R_o outer-edge radius. h thickness of cone material. E Young's modulus. ϱ density. ν Poisson's ratio. D internal loss factor. PC polycarbonate. CAB cellulose acetobutyrate.

Cone number	Geometry				Material				
	α (°)	R_i (mm)	R_o (mm)	h (mm)	E (10^9 N/m^2)	ϱ (kg/m^3)	ν	D	
50.1	50	17	83	0.23	2	600	0.3	0.1	
50.2e	50	17	83	0.23	2.4	1200	0.35	0.014	PC
50.3	50	17	83	0.23	2.2	1160	0.3	0.1	
50.3e	50	17	83	0.27	2.2	1160	0.34	0.06	CAB
60.1	60	17	83	0.1	2	600	0.3	0.1	
60.2	60	17	83	0.26	2.2	1160	0.3	0.1	
60.2e	60	17	83	0.26	2.2	1160	0.34	0.06	CAB

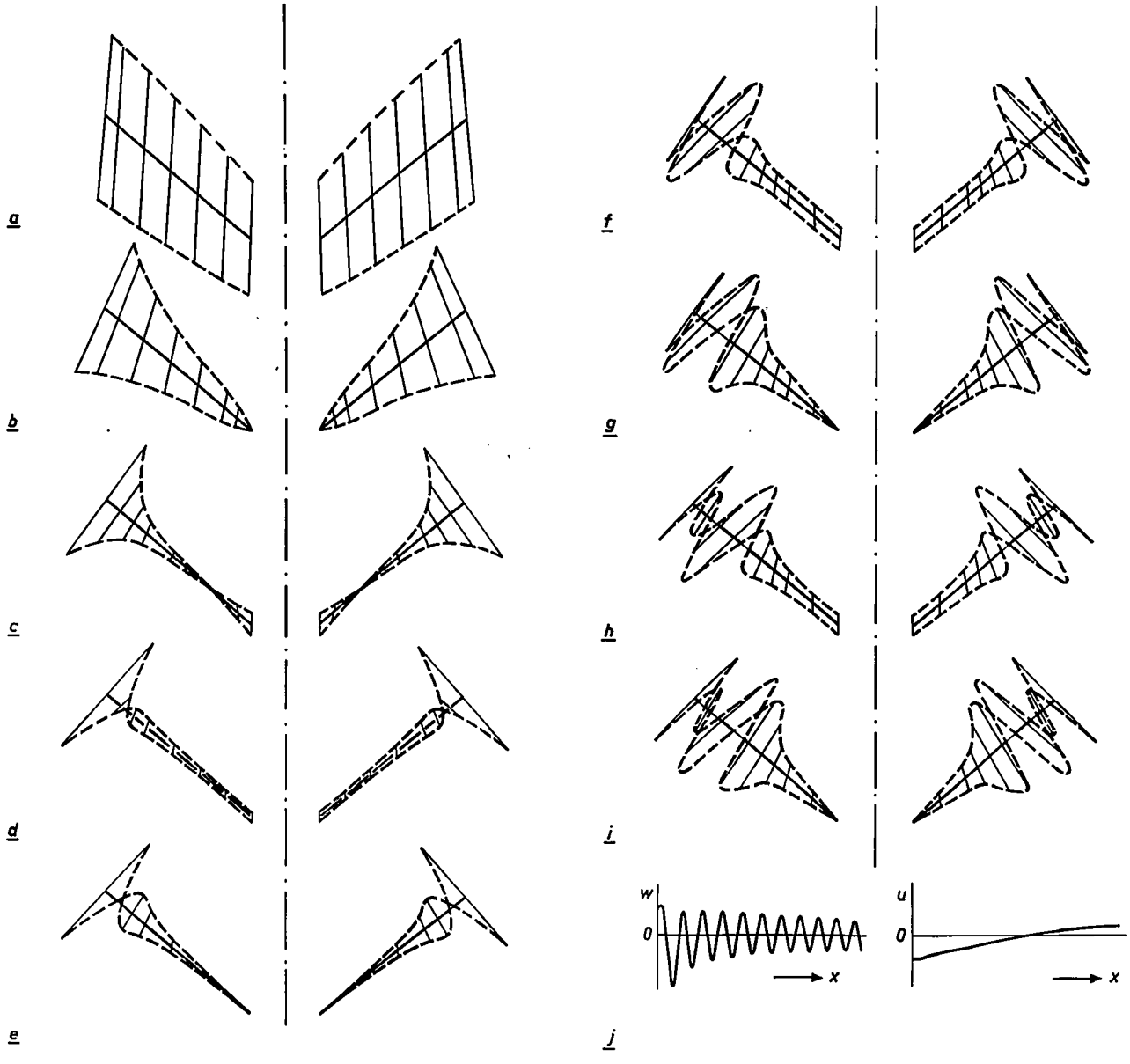


Fig. 9. Vibration patterns, composed of the calculated transverse and longitudinal displacements w and u on cone 50.1 (see Table I). The figures are not to the same scale.
a) 1000 Hz
b) ring antiresonance, $f_{ra} = 1840$ Hz
c) 2200 Hz
d) first bending resonance, $f_{br1} = 2360$ Hz
e) first bending antiresonance, $f_{ba1} = 2418$ Hz
f) second bending resonance, $f_{br2} = 2668$ Hz
g) second bending antiresonance, $f_{ba2} = 2750$ Hz
h) third bending resonance, $f_{br3} = 2993$ Hz
i) third bending antiresonance, $f_{ba3} = 3083$ Hz
j) 13 000 Hz; the curves of w and u along the cone are plotted separately

of the transition circle at the outer edge of the cone, i.e. up to frequency f_{to}. In this region there are as yet no bending waves on the cone. These first appear in region II, in which the ring antiresonance gradually progresses from the outer to the inner edge. In the innermost part the cone motion is still almost uniform. At the frequency f_{ti} the ring antiresonance has reached the inner edge; this is the upper limit of region II. In region III the ring antiresonance has disappeared from the cone and the entire surface of the cone is covered with bending waves.

Displacement patterns for frequency region I are given in fig. 9*a*, *b* and *c*. Fig. 9*a* shows the vibration behaviour at 1000 Hz. At this relatively low frequency the cone may be considered as a first approximation to vibrate as a rigid body. When the frequency is raised the amplitudes at the inner edge decrease whereas they increase at the outer edge. At the ring antiresonant

frequency $f_{ra} = 1870$ Hz the longitudinal amplitude becomes zero at the inner edge: a node appears. Because of the assumed rigidity of the inner edge, this node also appears in the transverse amplitude (fig. 9*b*). When the frequency is raised further, the node moves to the outer edge, as shown in fig. 9*c* (2200 Hz). The upper limit of region I lies at $f_{to} = 2250$ Hz.

Above f_{to} standing bending waves appear on the cone at certain frequencies; these can be divided into bending resonant frequencies, for which $u(a)$ and $w(a)$ go to infinity, and bending antiresonant frequencies, at which $u(a)$ and $w(a)$ become zero.

At the first bending resonant frequency $f_{br1} = 2360$ Hz (fig. 9*d*), w and u become very large. At the first bending antiresonant frequency $f_{ba1} = 2418$ Hz a new nodal circle appears in $w(x)$ at the inner edge (fig. 9*e*); this moves relatively quickly to the outer edge when the frequency is further increased.

The vibration patterns at the second and third bending resonant and antiresonant frequencies are given in fig. 9 *f-i*. The figures all show the same general picture: inside the transition circle the wavelength on the cone is long and the motion is approximately axial, because $u(x)$ and $w(x)$ have the same order of magnitude; outside it the vibration amplitude is mainly determined by the much shorter bending waves, which have a much greater amplitude than the longitudinal waves.

In frequency region III the coupling between bending and longitudinal waves becomes weak. The wavelength of the bending waves, which now cover the entire cone, is approximately equal to the bending wavelength on an infinite plate. In fig. 9*j* the transverse and longitudinal displacement patterns at $f = 13\,000$ Hz are of necessity shown separately, since in this case a combined picture would no longer give a clear picture of the vibration pattern.

Holographic display of vibration patterns

Minute displacements or deformations of an object can be made visible by means of holography [3]. We have used the time-averaged holography technique to render visible the vibration patterns of a loudspeaker cone to verify the results of the calculations qualitatively. We chose for the purpose a plastic cone (made of polycarbonate, cone 50.2e), since this material is more homogeneous than that of the more usual paper cone; the vibration patterns are consequently more regular and give a clearer picture. The geometry was the same as that of cone 50.1, but the constants of the material were of course different (see

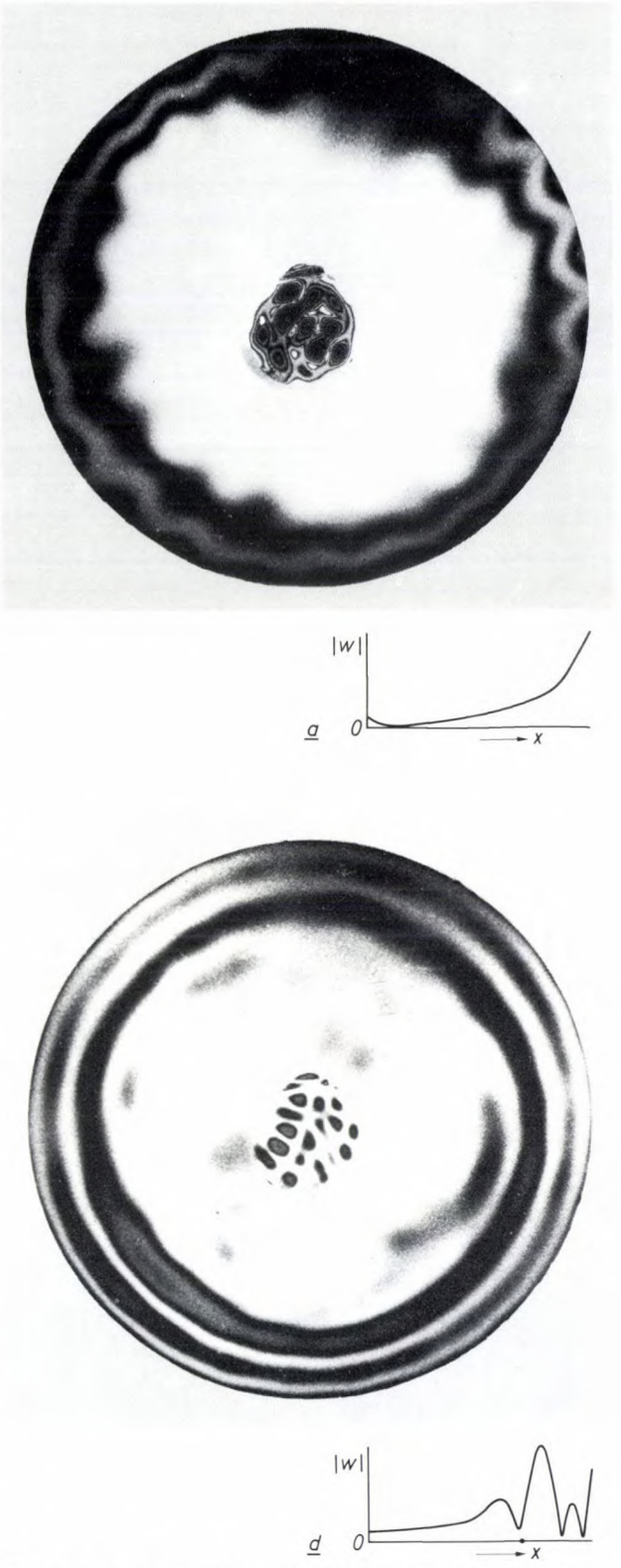
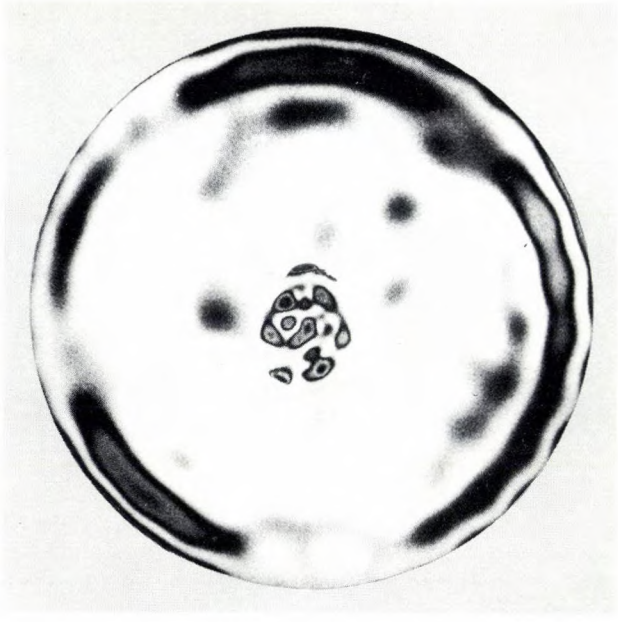
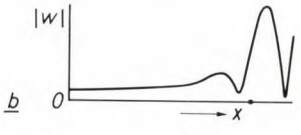
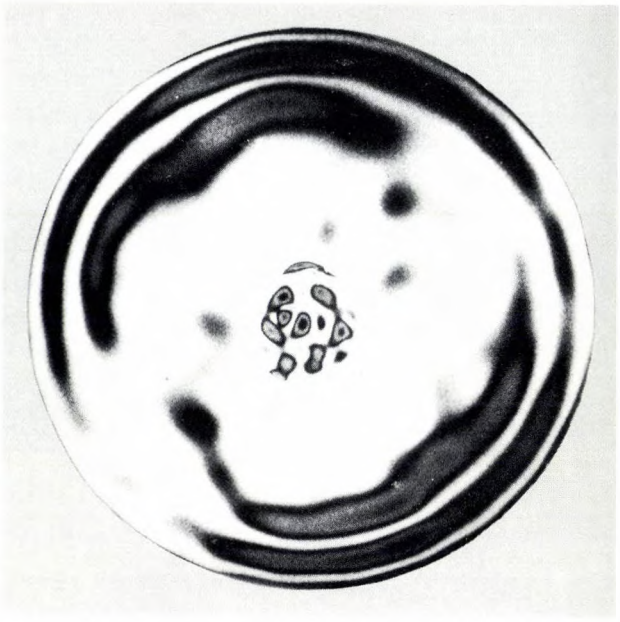


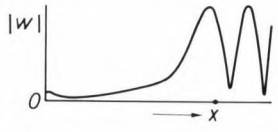
Fig. 10. Vibration patterns of a plastic cone (cone 50.2e), made visible by holography. White: low transverse amplitude (nodal lines), grey or black: high amplitude. The calculated amplitude curve is shown below each photograph ($x=0$ at the inner edge of the cone).
a) ring antiresonance, $f_{ra} = 1646$ Hz
b) second bending resonance, $f_{br2} = 2063$ Hz
c) second bending antiresonance, $f_{ba2} = 2170$ Hz
d) third bending resonance, $f_{br3} = 2337$ Hz
e) 6432 Hz
f) 8956 Hz (upper limit of region II $f_{ti} = 8520$ Hz)

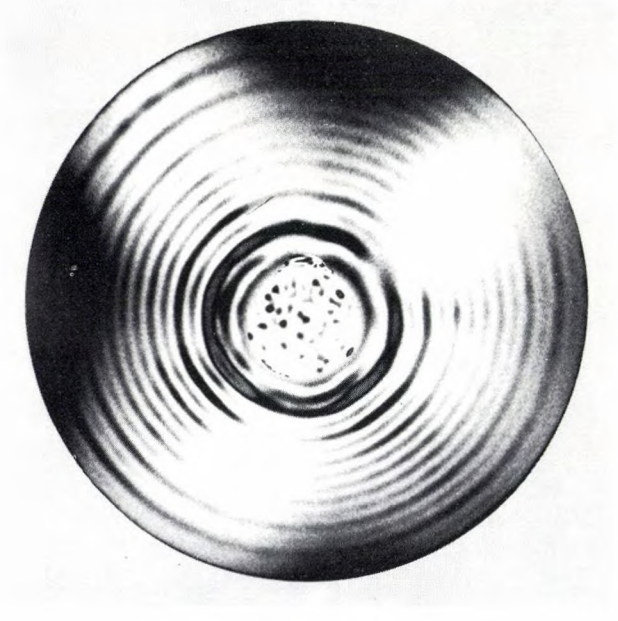
[3] See C. H. F. Velzel, Holographic strain analysis, Philips tech. Rev. **35**, 53-64, 1975 (No. 2/3).

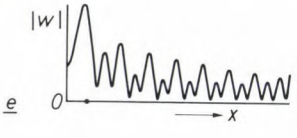


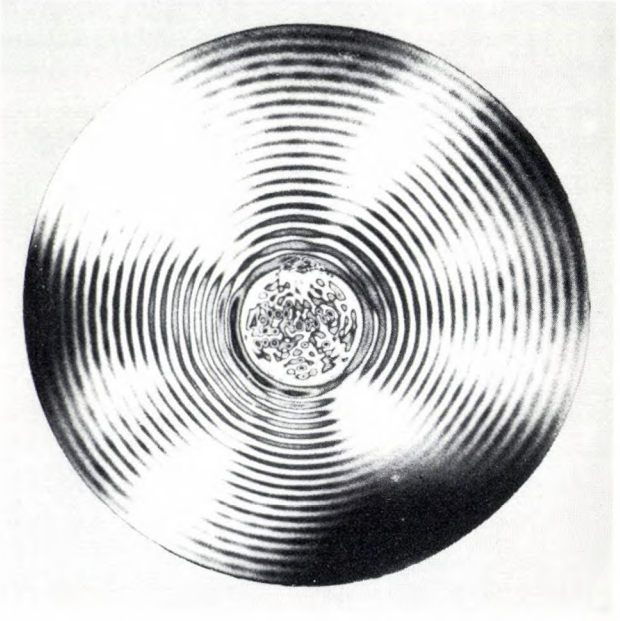

|w|
b
0
x




|w|
c
0
x




|w|
e
0
x



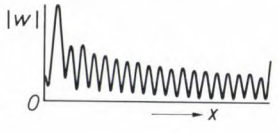

|w|
f
0
x

Table I). The recordings were made with the cone vibrating in air with a free outer edge.

The results are shown in *fig. 10a-f*. In the lighter parts the transverse amplitude is small (nodal lines), in the grey and black parts it is large. The calculated curve of the transverse amplitude $w(x)$ is shown beneath each photograph; damping was taken into account in the calculations.

Fig. 10*a* was recorded at the frequency f_{ra}. Besides the large amplitude at the outer edge, the picture shows a standing-wave pattern of circulating waves with ten nodal diameters. The grey ring near the edge is not a node but originates from the high local amplitudes. The patterns in fig. 10*b*, *c* and *d* were recorded at bending resonant and antiresonant frequencies; the transition circle is indicated in the graphs by a point. In fig. 10*e* (6432 Hz) the transition circle has almost reached the inner edge; in fig. 10*f* (8956 Hz, i.e. just above $f_{ti} = 8520$ Hz) it has disappeared from the cone, which is now completely covered by bending waves. The holographic pictures provide a complete confirmation of the calculated transverse-amplitude patterns.

The inhomogeneities on a paper cone are usually greater and cause a more distorted pattern. Additional distortions of the symmetry may be caused by the outer suspension. This is illustrated in *fig. 11*, which shows a holographic presentation of the vibration pattern of a commercial loudspeaker with a paper rim. These irregularities do not adversely affect the sound reproduction from a paper cone, but make such a cone less suitable for demonstrating the vibration patterns.

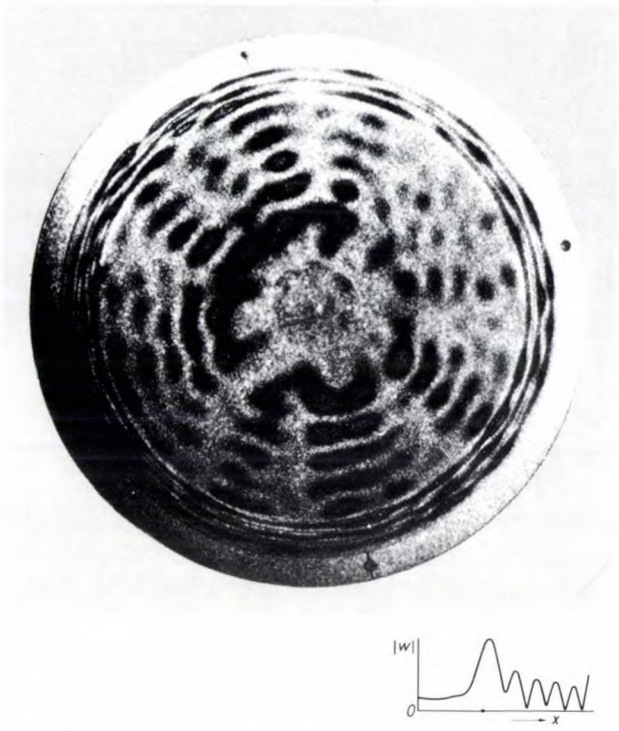


Fig. 11. Vibration pattern of a loudspeaker with paper cone (Philips 9710), made visible by holography. Frequency 5929 Hz. The pattern is less regular than that in fig. 10 because a paper cone is less homogeneous than a plastic cone.

Sound radiation from a flexible cone

The sound radiation from a flexible cone may be calculated in essentially the same way as that from a rigid piston and rigid cone. Each element of the cone surface is regarded as a point source and the contributions of all the point sources are added together. It is again assumed that the loudspeaker is enclosed in an infinite baffle. A complication here is that each surface element has its own transverse vibration amplitude, which must first be calculated and then included in the acoustic summation as a weighting factor.

Another complication is that, because of the succession of resonances and antiresonances in the flexible cone, the mechanical impedance to the motion of the voice coil varies considerably with frequency. With a constant drive force this means there is great variation in velocity amplitudes at the inner edge of the cone. This variation is also of course reflected in the freqency response; for a given drive force the sound power radiated at a given frequency is inversely proportional to the mechanical impedance presented to the voice coil, or, in other words, directly proportional to the mechanical admittance. This quantity therefore deserves to be looked at more closely.

Mechanical admittance

Like the vibration pattern, the mechanical admittance at the inner edge can also be calculated for any frequency from the set of differential equations mentioned in the previous section. As an example *fig. 12* shows the calculated curve of the axial admittance Y_a as a function of frequency for cones 50.1 and 60.1 (see Table I). At low frequencies, where the cone vibrates as a single mass, inertia is the decisive factor and the admittance decreases with increasing frequency. A minimum is reached at f_{ra}; just above it is the frequency f_{to} (indicated by a point), where the bending resonances and antiresonances begin. In cone 50.1 these give a ripple in the admittance curve. Above f_{ti} (the second point) the bending resonances and antiresonances are no longer perceptible in the admittance curve. The marked oscillations of the curve are connected with the longitudinal resonances and antiresonances.

Cone 60.1 differs from cone 50.1 in its larger apex angle and thinner paper. Because of the larger apex angle the characteristic frequencies f_{ra}, f_{to} and f_{ti} are lower. The thinner paper makes the cone lighter, so that the curve starts higher and the admittance varia-

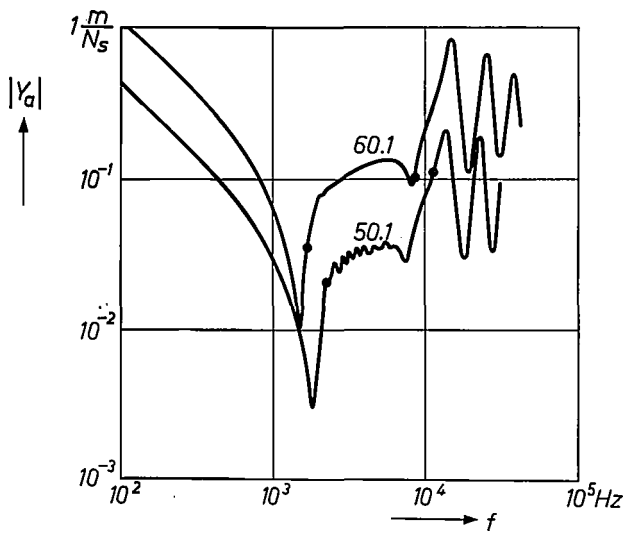


Fig. 12. Calculation of the axial mechanical admittance Y_a at the inner edge of cones 50.1 and 60.1 as a function of frequency. The 'points' indicate the locations of frequencies f_{to} and f_{ti}. At low frequencies the cone mass predominates. At f_{ra} a minimum occurs; f_{to} lies just above it. Between f_{to} and f_{ti} the bending resonances and antiresonances in cone 50.1 give rise to a fine structure. Above f_{ti} the curve is determined by the longitudinal resonances and antiresonances.

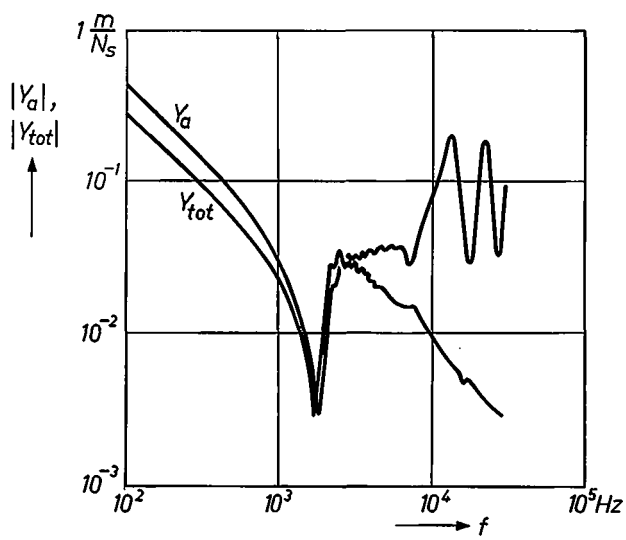


Fig. 13. Calculation of the total axial mechanical admittance Y_{tot} as a function of frequency after the addition of the voice-coil mass to cone 50.1. Curve Y_a from fig. 12 has been added for comparison.

tions associated with the bending waves are less pronounced; in fig. 12 they can no longer be seen.

The mass of the voice coil is not taken into account in fig. 12. In practice its effect is considerable and can even be dominant at high frequencies; the admittance oscillations due to the longitudinal resonance effects cause no more than a ripple on the curve determined by the voice-coil mass (see *fig. 13*).

Calculated frequency characteristics

To draw the frequency characteristic of a loudspeaker in sufficient detail it is necessary to perform calculations at some fifty well chosen frequencies. The calculation of localized vibration amplitudes takes a considerable amount of computer time (about two minutes for each frequency with an IBM 370/168; the acoustic variables take somewhat less). Our initial calculations were made with the mass of the voice coil taken as zero. This gives a basic curve, and very little extra computing time is required in correcting for the effect of the voice-coil mass, which is of a simple nature. This allows different values for this important parameter to be introduced at a later stage for a rapid determination of its effect on the curve.

Fig. 14 shows the result of calculations on cone 50.1. First of all we calculated the level L_p of the sound pressure at an axial distance of 10 metres from the loudspeaker (fig. 14*a*). A drive force with an amplitude of 1 N is assumed. We also calculated the level L_{P100} of the acoustic power radiated in a conical region of apex angle 100° (fig. 14*b*). Calculation of the sound pressure is unrealistic for field points lying outside this conical region, which is bounded by the loudspeaker cone itself with its apex angle of 100°.

The results may be compared with the solid curves in fig. 3, relating to a rigid cone of the same dimensions. Below f_{ra} (1840 Hz) the sound radiation from both the flexible and the rigid cone is almost identical with that from a rigid piston. Since the mechanical admittance has a minimum at f_{ra}, we also expect a minimum there

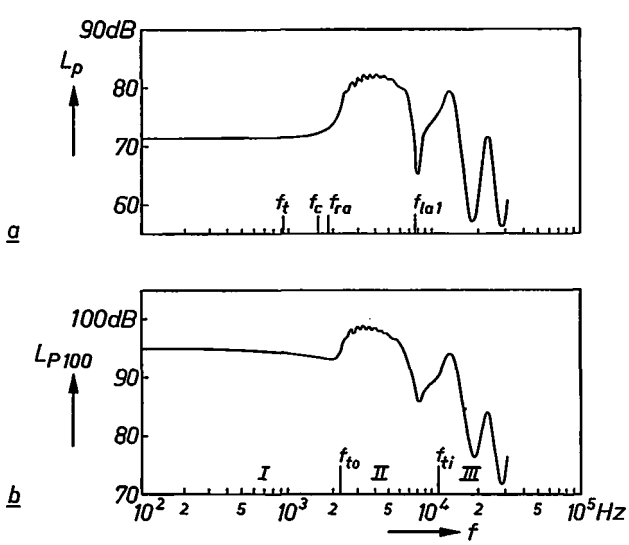


Fig. 14. *a*) Sound-pressure level L_p on the axis of cone 50.1 (calculated curve neglecting voice-coil mass; distance 10 m, drive force 1 N). The level is expressed in decibels relative to a reference level of 20 μPa. *b*) Acoustic power level L_{P100}, radiated within a conical region of apex angle 100°. The level L_{P100} is expressed in decibels relative to 10^{-12} W. Some characteristic frequencies and the frequency regions I, II and III are given.

in the sound-pressure response; this is not found because of the compensation provided by the high transverse amplitude at the outer edge (see fig. 10*a*). The decrease in the sound pressure expected for a cone above f_c (here 1580 Hz) cannot be seen because f_c and f_{ra} are relatively close together. The power response shows a slight decrease above f_t (920 Hz); see fig. 14*b*.

In frequency region II ($f_{to} < f < f_{ti}$) the sound radiation is predominantly controlled by the relatively uniformly vibrating inner part of the cone; the short bending waves on the outer part of the cone are 'acoustically short-circuited' and radiate little sound. The characteristics in this region have a broad maximum with a superimposed fine structure of bending resonances and antiresonances, followed by a deep minimum at the first longitudinal antiresonant frequency f_{la1} (7513 Hz), where there is a minimum in the mechanical admittance (fig. 12, curve 50.1). In the high-frequency region (region III, $f > f_{ti}$) the pressure response shows the same oscillating character as the mechanical admittance; peaks and dips alternate at longitudinal resonant and antiresonant frequencies.

If we now take the voice-coil mass into account, the characteristics change shape at the high frequencies. We have already noted that the mechanical admittance at high frequencies is entirely determined by the voice-coil mass; in the acoustic response curves this appears in a steep drop in frequency region III (see *fig. 15*). If this drop begins above the sharp minimum at the first longitudinal antiresonant frequency f_{la1}, this frequency will then in practice be the upper limit of the frequency range of the loudspeaker. The decrease due to the voice-coil mass, however, may begin in region II.

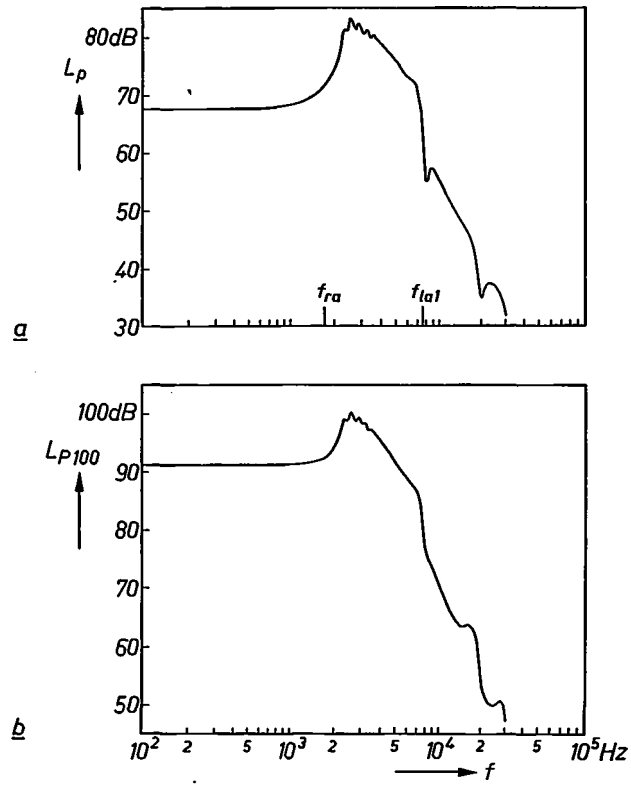


Fig. 15. *a*) Sound-pressure level L_p and *b*) power level L_{P100}, calculated for cone 50.1, with the voice-coil mass assumed equal to half the cone mass.

Comparison of measured and calculated characteristics

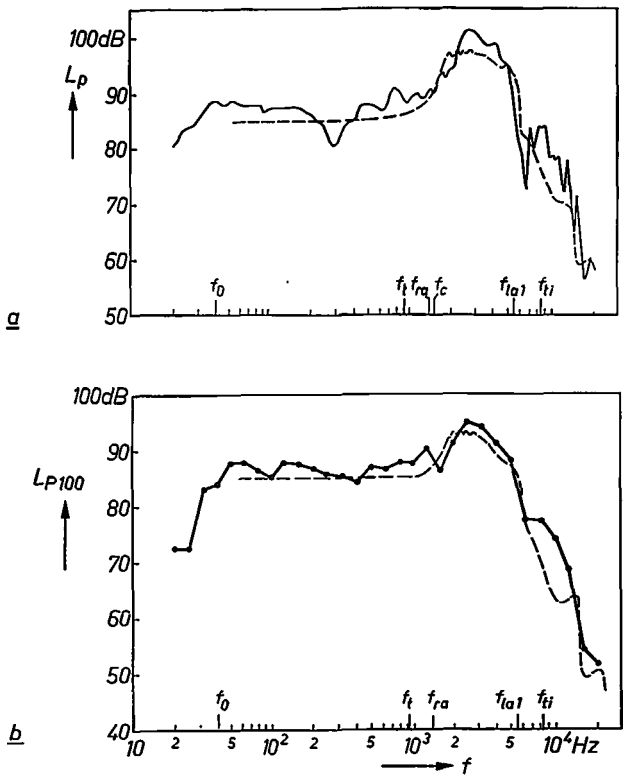
To test the theory against practical experience, the sound radiation was both measured and calculated for cones of various apex angles. Here again, plastic cones were used, but this time of different composition (cellulose acetobutyrate). The cones were provided with a rubber outer rim. The internal damping or loss factor of the material was 0.06, which is considerably more than that of the polycarbonate used for the holographic recordings. In those experiments it was important to make the bending waves visible, but in the acoustic measurements it was important to damp the bending resonances. A loss factor of 0.1 was assumed to allow for the damping at the rim and the radiation damping. The assumed infinite baffle was approximated by a closed acoustic box (volume 100 dm^3) attached to the middle of a square baffle (1.5×1.5 m^2).

In *fig. 16a* the solid curve is the measured sound-pressure response of a cone with a semi-apex angle of 50° (cone 50.3e, in Table I); the dashed curve is the calculated response for the same cone. The calculated values of the various characteristic frequencies are indicated. The main features of the measured curve can all be explained from these values. The maximum in frequency region II does indeed lie between f_{ra} and f_{ti} and the characteristic does start to fall off at f_{la1}.

The dip in the measured curve at 300 Hz is related to the dimensions of the baffle. At low frequencies the rest of the measured curve lies above the calculated one, because of radiation from the outer suspension. Above f_{ra} this suspension contributes little to the sound radiation.

Measurements were also made of the total sound power radiated within a conical region of apex angle 100°, with ten microphones arranged in an arc around the loudspeaker. Because the ten signals had to be summed, measurements could only be made at a number of discrete frequencies, yielding the values indicated by the points in fig. 16*b*. In fig. 16*c* the measured and calculated directivity diagrams are compared.

The characteristics relating to a cone with a semi-apex angle of 60° (cone 60.2e, Table I) are given in *fig. 17*. They give a lower maximum in region II. Although they do not differ fundamentally from fig. 16, most of the characteristic frequencies are lower.



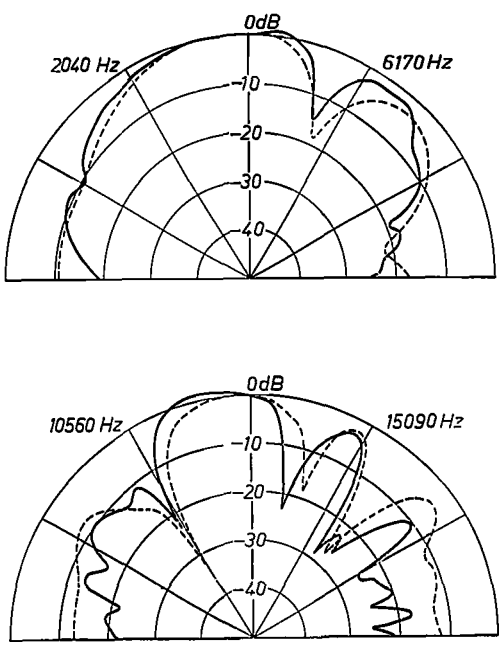


Fig. 16. Comparison of measured and calculated characteristics of cones 50.3e and 50.3; the dashed curve relates to cone 50.3. The calculated characteristic frequencies are given. *a*) Sound-pressure level. *b*) Level of total acoustic power radiated within a conical region of apex angle 100°. *c*) Directivity characteristics. The sound-pressure level is plotted as a function of the direction of radiation.

It may be concluded from the good agreement between the calculated and measured characteristics that the calculation procedure gives a good approximation to the actual behaviour of the cone.

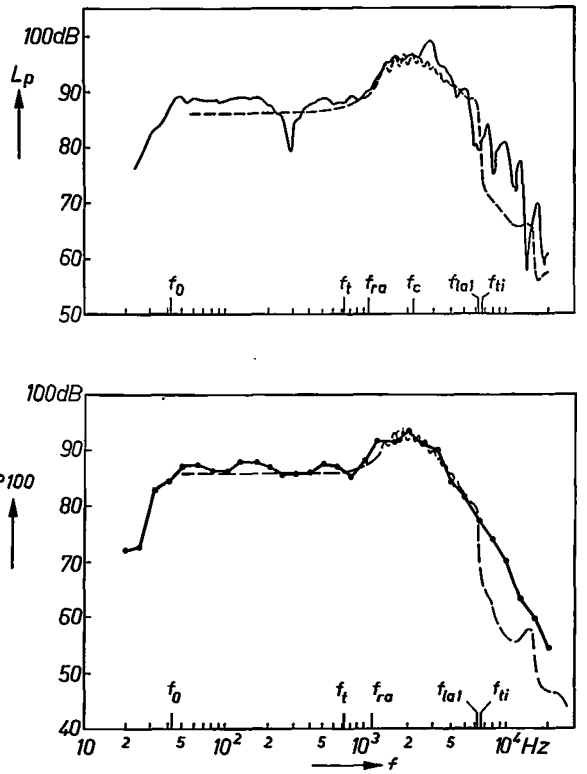


Fig. 17. Comparison of measured and calculated frequency characteristics of cones 60.2e and 60.2.

The design of a loudspeaker cone

The primary consideration in the design of a loudspeaker cone is to ensure that the characteristic frequencies are properly located.

All the requirements cannot be satisfied simultaneously. A flat pressure and power response, a large bandwidth and a high efficiency cannot all be achieved at the same time. Every design must therefore be a compromise.

If the flattest possible characteristics are required, then to avoid the dip at f_{ra} in the power response, f_{ra} should not lie too far above f_t. This has consequences for the apex angle. If, for example, we take f_{ra} smaller than $2f_t$, the semi-apex angle α of a paper cone, in which the propagation velocity c for longitudinal waves is 2700 m/s, must be greater than 70°. Such a large value of α implies a low maximum in region II, which will help to give a flat response. In this region, however, a distinct fine structure may easily arise, because with a large apex angle the frequencies of the bending resonances and antiresonances lie further apart; α should therefore not be made larger than is absolutely necessary.

The steep decrease at high frequencies, caused by the voice-coil mass, generally determines the upper limit of the frequency range of the loudspeaker. To achieve a high cut-off frequency the ratio m_c/m_d between the

masses of voice coil and diaphragm must be chosen as small as possible. This can never be higher, however, than f_{la1} or f_{ti}, whichever of the two is lower, since the cone behaviour sets an upper limit here to the frequency response. The frequencies f_{la1} and f_{ti} depend on the radius of the inner edge; if the radius is smaller, f_{ti} is higher but f_{la1} lower.

The cone thickness h can be determined experimentally after the other geometrical parameters have been determined. A thin cone helps to suppress the fine structure of the bending waves but makes the ratio m_c/m_d worse. It also increases the danger of nonlinear distortion because the amplitudes of vibration become too large. One means of eliminating fine structure is to use a damping outer rim.

If the primary requirement is a large bandwidth, then as noted above, the ratio m_c/m_d must be given the smallest possible value. This implies a heavy cone and low efficiency. If dips in the response at f_{ra} and f_{la1} are acceptable, a small value of α should be taken and c made as high as possible; f_{ra}, f_{la1} and f_{ti} then become higher. A semi-apex angle of 50°, for example, can then be used, which makes the maximum in region II relatively high. The thickness is again chosen in such a way as to eliminate the fine structure.

If high efficiency is most important, the cone should be made as light as possible. The ratio m_c/m_d then becomes higher, however, and the bandwidth will consequently be small. A high maximum in region II can be obtained by choosing a small apex angle. If a dip in the pressure response at f_{ra} is not allowed, the Young's modulus of the cone material should be chosen such that f_{ra} is lower than $2f_c$.

Computer-aided design with visual display

It would be ideal if the designer could feed a sketch of the desired frequency characteristic into a computer and get back a little while later the corresponding dimensions and material constants for the cone. This is not feasible, however, since not all the curves can be realized in practice. Computer-aided design of a loudspeaker cone still has to be done the other way round: the designer types in some data and the computer calculates the resultant frequency characteristic. If it differs from the required curve, the designer then changes one or more parameters and the computer presents another result. An iterative procedure of this type is best carried out at a computer terminal with a visual display (*fig. 18*).

Such a procedure requires a great deal of computer time: it can take an IBM 370/168 computer as long as an hour to calculate a single characteristic. If the procedure is to be used frequently, costs can be reduced by storing a number of standard characteristics in the computer memory and using these as the starting point for each new design. These characteristics can be adequately characterized by four parameters: the outer radius R_o, the inner radius R_i, the semi-apex angle α and the velocity c of longitudinal wave propagation in the cone material.

Poisson's ratio (of lateral to longitudinal strain) has little influence, and is set at $\frac{1}{3}$ in all cases. The loss factor giving the internal damping of the cone material is taken at the relatively high value of 0.1, and the cone thickness is given the empirical value $R_o/800$; both values reflect the practical measures that need to be taken to avoid a fine structure.

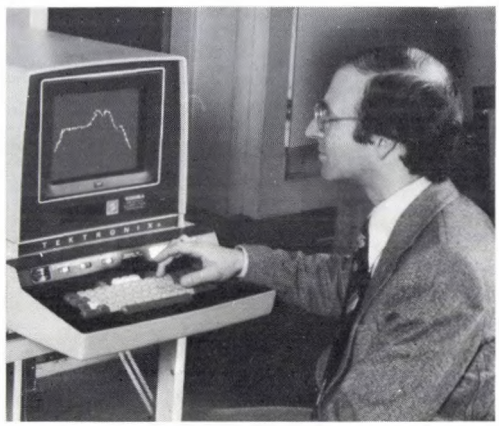


Fig. 18. Computer-aided design of a loudspeaker cone using a visual display.

For R_o, for example, seven values can be taken (corresponding to the commercial diameters 3, 4, 5, 6½, 8, 10 and 12 inches), for R_i three values (e.g. 18, 35 and 50 mm), for α five values (50° to 70° in steps of 5°) and for c a single value (2500 m/s). The vibration patterns and frequency characteristics of 105 cones now have to be calculated and stored. From the stored vibration patterns the vibration patterns and frequency characteristics for other materials can be computed fairly quickly. If we take, for example, $c = 1500$ m/s and $c = 3500$ m/s, we can calculate the data for 315 cones and store them in the computer memory.

The designer can call up the characteristics of any one of these cones on his visual display. Having done this he types in a value for the ratio m_c/m_d of voice-coil mass to cone mass. The computer immediately calculates the effect of this ratio on the characteristics, which takes very little computer time.

If the desired characteristic is not sufficiently approximated, the designer can request calculations for intermediate values of the four basic variables; the programs for this are stored in the computer. The results are stored in the memory, so that the range of choices available is gradually widened.

After a trial model has been made in this way, it may be found that the characteristics have an undesired fine structure. The damping must then be increased or the cone thickness reduced; neither of these measures has much effect on the general shape of the frequency characteristics provided the ratio m_c/m_d and the fundamental resonant frequency f_0 are kept constant.

The spring constant of the cone suspension and the characteristics of the electrodynamic drive can be included in the model with very little extra complication. Their influence can be immediately calculated by the computer from relatively simple formulae. In this way a very comprehensive loudspeaker design can be produced.

Summary. A loudspeaker cone gives complex and highly frequency-dependent vibration patterns. With a computer numerical solutions can be found for the set of simultaneous differential equations that describe the vibration behaviour of a flexible conical shell. Three frequency regions are distinguished: I. low frequencies, the cone vibrates as a rigid body; II. a ring on the cone gives a special resonance, bending waves occur outside this ring; III. high frequencies, the entire cone is covered with bending waves. For the frequency characteristics of the pressure response and total sound-power response of the cone to have the desired shape, it is necessary for the boundaries between the regions to be correctly located; this determines the choice of the radii of the outer edge and the voice coil, the apex angle and the material constants of the cone. The bending resonances are less pronounced for a thinner cone or material with greater damping. For a high cut-off frequency the mass of the voice coil must be low compared with that of the cone. Computer-aided design is possible by calling up a visual display of previously stored calculated characteristics and modifying the parameters.

An easily controlled alkali-vapour dispenser

The photocathode of devices such as image amplifiers are often formed from an alkali antimonide. Coatings of these materials cannot be formed directly by vapour deposition of the material itself, however, since the material would dissociate in the process. Photocathode coatings are therefore made by means of separate sources that alternately deliver accurately controlled amounts of antimony and alkali vapour. The usual practice is to incorporate these sources, called 'dispensers', in the tube before evacuating it and sealing it off, and to produce the evaporation from outside by an electrical method. A computer is often used to determine the times at which the vapour production is to be switched from one source to another; it does this from data relating to properties such as the light transmission and the electrical resistance of the coating, which are continuously measured during manufacture. To avoid difficulties in manufacture it is important that the vapour production should be constant, especially after the temporary interruptions inherent in the production method. In constancy and degree of control it is the alkali dispensers that leave most to be desired.

Fig. 1 shows a cross-section and a photograph of a conventional type of sodium dispenser. The housing consists of a rolled-up strip of nickel-chromium steel. The filling may be a powder mixture of Na_2CrO_4 and zirconium. An electric current generates heat which produces the following reaction:

$$Na_2CrO_4 + 2Zr \rightarrow 2ZrO_2 + Cr + 2Na\uparrow.$$

The sodium vapour is released from the join or seam along the whole length of the tube.

Exposure of the alkali metal to air can be minimized by arranging for the sodium vapour to be released, by means of a chemical reaction, only at the exact time and location required for manufacture.

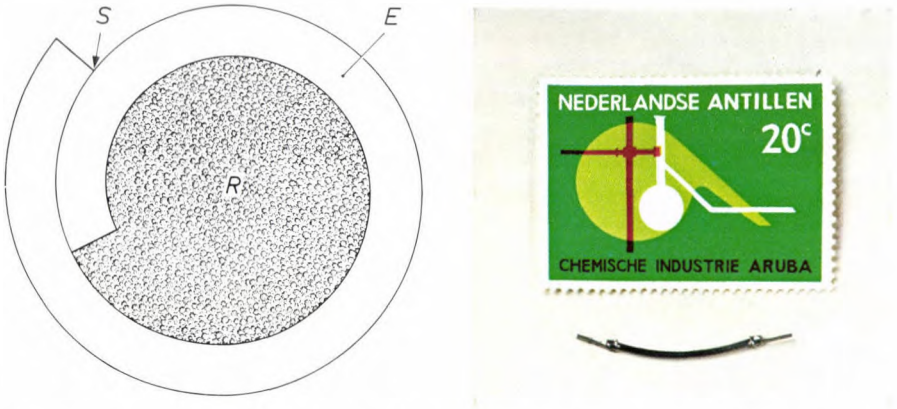


Fig. 1. Cross-section (left) and photograph of a sodium-vapour dispenser. Dispensers of this type are used for making photosensitive layers of alkali antimonide. The 'housing' *E* consists of a rolled-up strip of nickel-chromium steel (0.08 mm thick). The dispenser is connected to a constant-current source by means of two contacts (see photograph). The filling of conventional types of dispenser is a mixture *R* of Na_2CrO_4 and zirconium. When the dispenser is electrically heated, this mixture gives off sodium vapour, which is released from the seam *S* along the whole length of the tube.

Since the reaction in the chromate dispenser described here is exothermic, however, exact regulation of the sodium-vapour production is difficult. Another difficulty is that the vapour production is not completely homogeneous, because there are always concentration fluctuations in the powder mix, however thoroughly mixed, and these give rise to local variations in heat production.

We have now developed a dispenser whose operation is based on an endothermic reaction instead of an exothermic one. The sodium vapour is produced by the dissociation reaction

$$Au_2Na \rightarrow 2Au + Na\uparrow.$$

Au_2Na is an intermetallic compound that does not decompose in air, is brittle and is thus easily pulverized. The compound is made by melting a stoichiometric

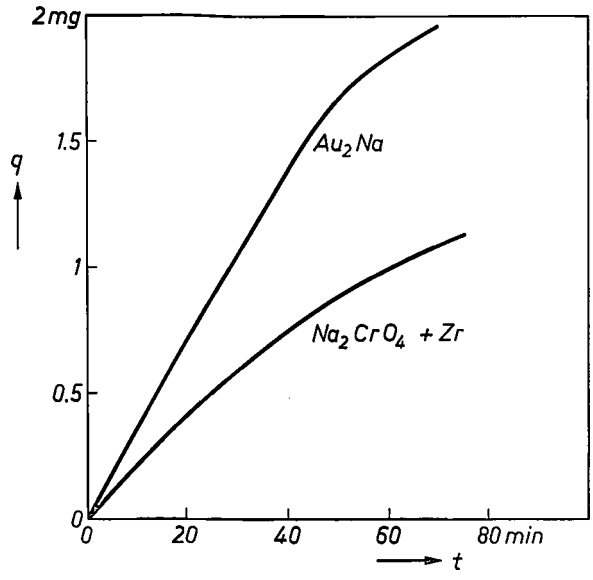


Fig. 2. The quantity q of sodium released as a function of time t at a current of 4.5 A for a sodium dispenser filled with Au_2Na (the new type) and for a dispenser filled with Na_2CrO_4 and Zr (the conventional type).

mixture of gold and sodium in a hermetically sealed molybdenum vessel. After the mixture has been pulverized to produce grains small enough to pass through a sieve with a mesh of 125 μm, the material is introduced into the same kind of container as the dispenser mentioned above.

Since the filling of the new dispenser consists of only one substance, the sodium vapour pressure it gives under equilibrium conditions depends only on temperature within very wide limits of composition as the reaction proceeds. In the earlier type of dispenser described above, the filling is formed from two substances, no equilibrium is established and the vapour production is also adversely affected by changes in the concentrations during operation.

The temperature at which the new type of dispenser delivers sodium is very much lower: 685 °C instead of 900 °C in the earlier type. The output is therefore greater for the same current (*fig. 2*). The vapour production of the new dispenser is also much less sensitive to interruptions. If the current increases by 0.1 A/min the first vapour appears at 4.1 A from the as yet unused chromate dispenser of the conventional type, but after the first interruption and cooling to room temperature the next release is found at 3.5 A. When the same procedure was followed with the new type, no change was observed: the current remained unchanged at 3.0 A. The amount of gas adsorbed from the air (H_2O, H_2, CO, CO_2 etc.), which is desorbed during the preliminary degassing, and also the unavoidable residue released during the vapour dispensing itself, are about 50% lower in the new type. Finally, the new type is more resistant to damp air: it can be kept for weeks, instead of days for the old type.

The advantages summarized here will be of obvious interest in automated production.

Potassium- and caesium-vapour dispensers based on the same principle are under development.

J. J. B. Fransen
J. H. N. van Vucht

Ing. J. J. B. Fransen is with Philips Elcoma Division, Eindhoven; Dr J. H. N. van Vucht is with Philips Research Laboratories, Eindhoven.

Observations of domains in ferroelectrics and ferromagnetics with a scanning electron microscope

C. Michel

The case described in this article — work with a scanning electron microscope — possesses two contrasting aspects. It shows, for both ferroelectricity and the electron microscope, that a scientific and technical subject can reach a stage of renaissance even without ever having been a true classic. For ferromagnetism, however, exactly the reverse is true. The subject really is a classic, but SEM observations by means of magnetic-contrast do not appear to be a very hopeful prospect for the future.

Detection of surface potentials

The scanning electron microscope offers about a dozen different ways of obtaining contrast in the images it can produce [1]. One of these is image formation by voltage contrast. In this method the secondary electrons emitted by the sample under investigation are detected. These secondary electrons are produced by the scanning (primary) beam of the SEM. The number of secondary electrons reaching the detector will vary with the electrostatic potential at the surface of the sample, giving a corresponding contrast in the image. This method of imaging has been widely used in the study of semiconductors and of microelectronic circuits [2].

The method can also be used to give a directly visible presentation of the domain structure in ferroelectric materials. This application is important because useful electro-optic phenomena such as electrically controlled birefringence and light scattering have been shown to be related to the ferroelectric domain configuration in the material concerned [3]. The study of the behaviour of ferroelectric domains — in which the object is to find ways of controlling them — is by no means easy. Certain interactions, such as those between electrical and mechanical quantities, make the treatment of ferroelectrics more complex than that of ferromagnetics.

In a ferroelectric the electrical polarization induces surface charges. These charges produce the potential fields on which image formation by voltage contrast depends. The various domains in general have different polarization directions, giving sufficient contrast for direct study of the domains with almost zero delay. Observation of domain structure by SEM does not necessitate the destruction of the material, as in techniques such as chemical surface etching [4].

The resolution of voltage-contrast imaging is in principle only limited by the diameter of the electron beam (0.02 μm). This value represents a considerable improvement: the conventional direct methods such as polarization microscopy and X-ray topography have a resolution of about 1 μm, because they depend on the use of an ordinary optical microscope.

It is also possible to use the SEM for a kind of 'dynamic' observation of ferroelectric materials. The electron beam then gradually builds up a surface charge that induces 'domain flipping'. This induced polarization switching and the associated redistribution of the ferroelectric domains can be observed with the same high resolution while it is actually taking place.

In the following sections some of our work in ferroelectricity, with both single-crystal and polycrystalline (ceramic) materials, will be discussed. The experiments clearly illustrate the feasibility of voltage-contrast imaging. A few preliminary details are given of the samples used and of the instrument settings. The final section of the article contains material on the — rather limited — possibilities of observing ferromagnetic domains by the somewhat analogous method of magnetic-contrast imaging.

Dr C. Michel, formerly with Philips Laboratories, Briarcliff Manor, N.Y., is now with Stauffer Eastern Research Center, Dobbs Ferry, N.Y., U.S.A.

[1] See for example D. B. Holt, M. D. Muir, P. R. Grant and I. M. Boswarva (eds), Quantitative scanning electron microscopy, Academic Press, London 1974; L. Reimer and G. Pfefferkorn, Raster-Elektronenmikroskopie, Springer, Berlin 1973; and W. Kuypers and J. C. Tiemeijer, The Philips PSEM 500 scanning electron microscope, Philips tech. Rev. 35, 153-165, 1975 (No. 6).

[2] P. R. Thornton, Scanning electron microscopy, Chapman and Hall, London 1968. See also W. H. Hackett, Jr, R. H. Saul, R. W. Dixon and G. W. Kammlott, J. appl. Phys. 43, 2857, 1972.

[3] C. E. Land, P. D. Thacher and G. H. Haertling, Electrooptic ceramics, in: R. Wolfe (ed.), Applied solid state science 4, 137-233, Academic Press, New York 1974.

[4] J. A. Hooton and W. J. Merz, Phys. Rev. 98, 409, 1955.

Equipment and materials

The four ferroelectrics that we investigated are shown in *Table I* with some of their characteristic data. The values for the spontaneous polarization apply in the region of room temperature. The first three compounds were investigated in the single-crystal form. The fourth compound, lanthanum-doped lead zirconate-titanate was the ceramic material studied. This material contains two crystal structures: α, which is tetragonal, and β, which is rhombic. Both depend on the thermal and electrical history of the sample.

Table I. The ferroelectrics that were investigated in a scanning electron microscope by means of voltage contrast. The column 'domain-wall type' gives the angle between the polarization vectors of adjacent domains.

Compound	Crystal class	Density (g/cm^3)	Spontaneous polarization ($\mu C/cm^2$)	Domain-wall type	Transition temperature (°C)
Triglycine sulfate $(CH_2NH_2COOH)_3.H_2SO_4$	Monoclinic	1.69	2.8	180°	49
Barium titanate $BaTiO_3$	Tetragonal	6.4	26	180°, 90°, 45°	130
Tetrabismuth-trititanium-12 oxide $Bi_4Ti_3O_{12}$	Monoclinic	6.6	4	180°, 90°	626
Lanthanum-doped lead zirconate-titanate 8/65/35 PLZT	Tetragonal (α) and rhombohedral (β)	7.6	28	properties vary with the α and β phase [9]	

The SEM used for the experiments was a standard production instrument whose voltage-contrast signal could be displayed on a television monitor. Images could also be recorded on a video recorder. With samples of soft material such as triglycine sulfate the SEM was operated at a low beam intensity (10^{-11} to 10^{-10} A) and a low accelerating voltage (2 to 6 kV). The surface then remained undamaged by the incident electrons.

In the 'dynamic' mode of observation a beam intensity of as much as 10^{-9} A was permitted. Under such conditions the electric field from the surface charge generated by the primary beam can easily reach several kilovolts per cm, at least for non-conducting materials. However, if some *existing* domain structure is being investigated — by observation in the static mode — the generation of static charge must be avoided. Non-conducting ferroelectrics must then be coated with a conducting film. Our samples were coated with a carbon film of thickness between 0.01 and 0.05 μm. Such a film has no significant effect on the resolution of the SEM. Thicker films could not be used because they would give too much screening of the electric field — responsible for the contrast — at the surface of the sample.

Static imaging of domains by voltage contrast

The first ferroelectric sample observed by the voltage-contrast (VC) technique was an uncoated optically flat surface of a single crystal of triglycine sulfate. A domain pattern is clearly visible in *fig. 1*. The domains — which are elongated along a preferential direction — are found to be separated by 180° walls, i.e. the adjacent domains



Fig. 1. The surface of a single crystal of the ferroelectric TGS (triglycine sulfate). Adjacent domains have opposite polarizations. The image, obtained with a scanning electron microscope, is produced by the effect of the electric field from the various domains on the secondary electrons emitted by the surface ('voltage contrast'). The surface of the sample was optically flat and uncoated. The small white spots denote the build-up of surface charge.

have opposite polarizations. The fact that 180° walls are easily observable represents a specific advantage of using the SEM. The conventional method of observation, with an optical microscope and polarized light, will not normally make such walls directly visible [5].

To confirm that fig. 1 is indeed due to voltage-contrast imaging we made two control experiments. Firstly we investigated the same uncoated samples in the SEM again, by using back-scattered electrons. However, no contrast could be seen. This control experiment proves that the pattern of fig. 1 cannot be due to topographical features of the surface. We also found that an uncoated sample with a screening gold layer of excessive thickness (more than 0.1 μm) did not produce any secondary-electron imaging. From these two experiments it can be safely concluded that fig. 1 does indeed represent a voltage-contrast image.

The domain patterns observed were found not to be particularly stable. They included white spots, which are characteristic of the build-up of surface charge.

After the triglycine-sulfate crystals we examined barium titanate. These crystals are much harder and easily withstand primary beams up to 10^{-10} A at accelerating voltages to a maximum of 24 kV. *Fig. 2* shows domains with 90° walls, delineated on a cleaved section of the crystal.

Fig. 3 shows a cleavage section of the third kind of single crystal: $Bi_4Ti_3O_{12}$. The typical striped structure is similar to that of fig. 2, although the domain configuration is more complex, owing to the rather unusual crystallographic structure of the material. In the crystal of fig. 3, which had the shape of a wafer, the direction of the spontaneous polarization of all domains was parallel to the surface of the sample. This meant that there was no contrast at the upper surface of the sample, which is consistent with the expected constant surface potential there.

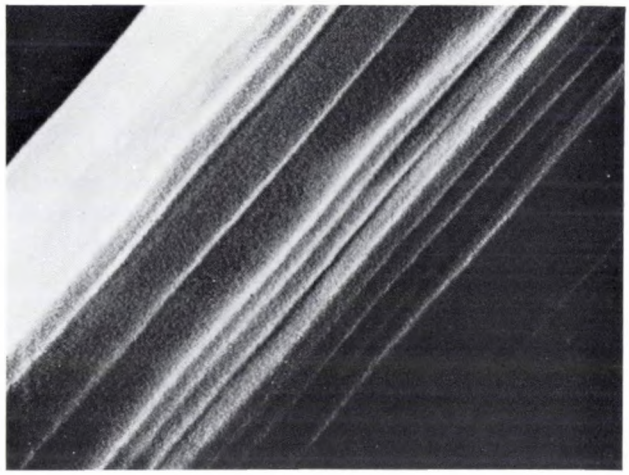


Fig. 2. Domain structure of an uncoated cleavage surface of a single crystal of $BaTiO_3$. The polarization vectors in adjacent domains are perpendicular to each other. The image was made with a scanning electron microscope (SEM), using voltage contrast.

Dynamic imaging by voltage contrast

Beam-induced polarization switching

The domain flipping in ferroelectrics and its interaction with the crystal properties is a complex phenomenon [6]. It is known, for example, that when the macroscopic polarization of a ferroelectric crystal is reversed by an external electric field, the domain configuration undergoes drastic changes. To obtain an understanding of the process much work has been done on various compounds, including the material we selected, triglycine sulfate. Using the SEM we were able to directly observe beam-induced domain flipping in an uncoated triglycine-sulfate crystal by means of the television monitor coupled to the microscope. The thickness of the wafer was about 50 μm, and as stated earlier the primary electron beam had an intensity of up to 10^{-9} A. At these current levels the observed domain pattern was found to be unstable and to be extremely sensitive to the beam. Granular white spots would suddenly appear on the monitor screen, indicating a build-up of surface charges.

When a strong field, of the order of several kilovolts, is present, the polarization switches very rapidly — in a

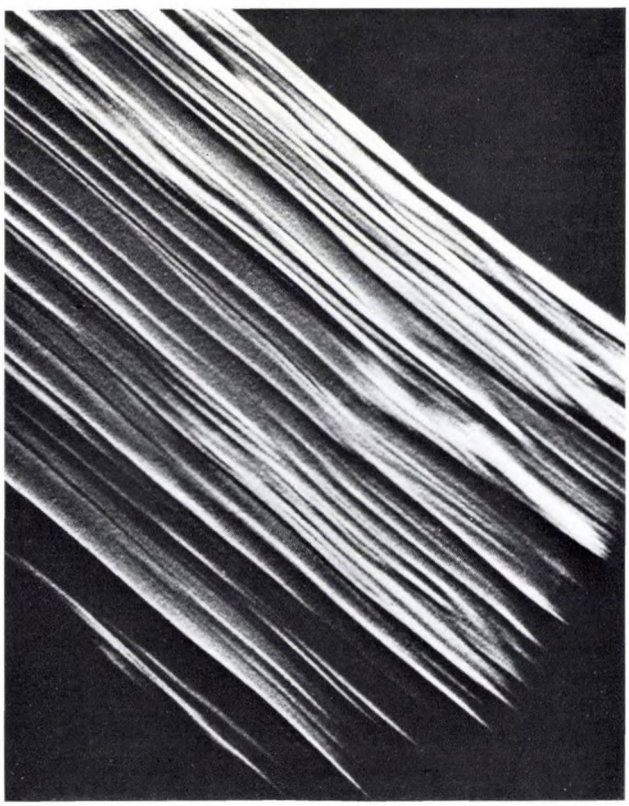
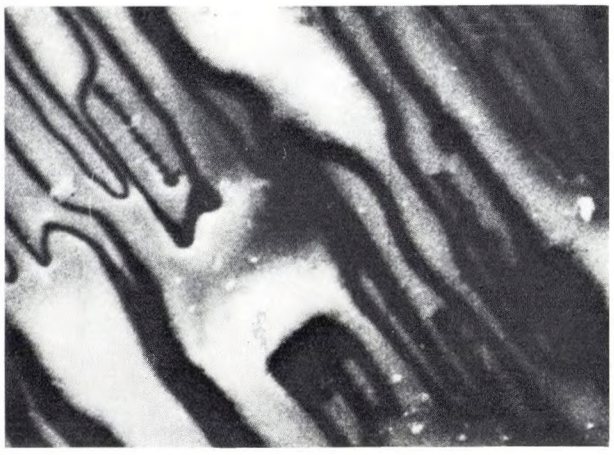


Fig. 3. Domain structure of an uncoated cleavage surface of a single crystal of $Bi_4Ti_3O_{12}$.



a



b



c

Fig. 4. Domain imaging in the voltage-contrast 'dynamic' mode. It is called 'dynamic' because the scanning beam induces domain flipping. The sample is an uncoated single crystal of TGS. The pictures are observed without any significant time delay on a television monitor coupled to the SEM. *a*) The incident primary electrons rapidly neutralize the existing surface charge. After a few scans the contrast between adjacent domains (with 180° walls) has decreased to such an extent that only the domain walls are still clearly delineated. *b*) After a few more scans the induced fields become sufficiently strong to reverse individual polarization vectors; for example, a domain near the center has been switched and now appears as a dark island. *c*) After switching charge compensation occurs and the contrast disappears again. Note the increased domain size.

few microseconds. The velocities at which the walls move across the wafer are of the order of meters per second. From this it follows that the switching of an individual domain takes place too quickly for it to be observed directly. The time necessary for switching the polarization of the *entire* surface scanned by the beam can be controlled within a certain range by varying the setting of the SEM: the time increases when the scanned area is increased or the intensity of the beam is reduced.

Fig. 4 shows various stages of the switching process occurring at the crystal surface: these images were observed on the monitor screen without any significant delay. It can be seen that initially the contrast between adjacent domains with 180° walls decreased rapidly. Fig. 4*a* shows an example of such flat contrast: only the domain walls are clearly delineated. The rapid reduction in the contrast results from the fast neutralization of the surface charges originally present by the electron beam. After the sample has been scanned a few times the applied charge is large enough for its field to reverse the polarization. This switching process creates the dark islands shown in fig. 4*b*. Somewhat later charge compensation occurs and the contrast disappears again,

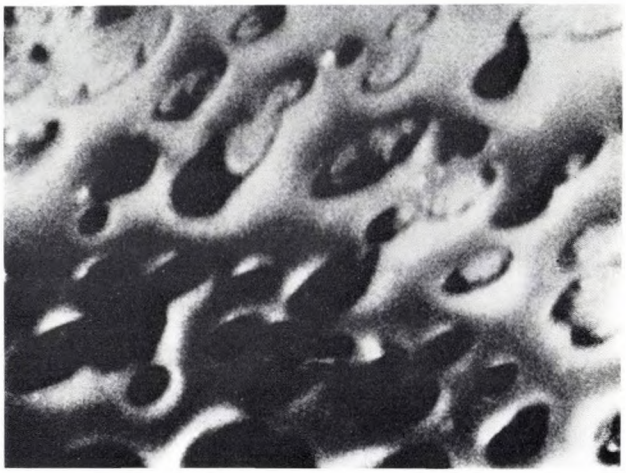


Fig. 5. An intermediate stage in the charge-compensation process (see fig. 4*c*). The areas inside the domain walls are still partly black, indicating that charge compensation is not yet complete.

leaving domains of larger size (fig. 4*c*). *Fig. 5* illustrates another intermediate stage of the charge-compensation process: it can clearly be seen that the process is only partially completed [7]. The contrast between domains of opposite polarization lasted for no longer than a few

[5] See for example F. Jona and G. Shirane, Ferroelectric crystals, Pergamon Press, Oxford 1962.

[6] See for example J. R. Maldonado and A. H. Meitzler, IEEE Trans. **ED-17**, 148, 1970.

[7] A discussion of charging effects is given in Thornton's book (see note [2] above), p. 111 and following.



Fig. 6. Photomicrograph of a single crystal of TGS, showing 'beam writing' produced by an SEM. The 'writing' consists of a row of dashes formed in non-adjacent domains of the same polarization. The dashes represent local reversals brought about by individual scans. The beam current was 10^{-9} A.

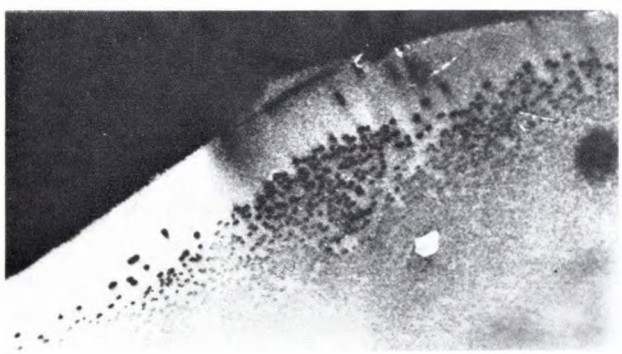
scans in this experiment; this period was so short that the contrast could only be observed in slow-motion playback on the video recorder.

By increasing the beam current to 10^{-9} A we were able to observe local switching induced by the individual scans (*fig. 6*). The parallel lines inside the domains indicate the individual scans. The lines are of course only visible in domains that have a polarization of the same sign, which means that they cannot be adjacent. Such a technique of 'beam writing' makes the SEM a convenient instrument for characterizing the polarization states and the domain structure at the surface of a ferroelectric sample.

The domain structure; nucleation effects

If high beam currents are used, local heating arising during the scan may be sufficient to cause the transition temperature of the ferroelectric material to be exceeded. In such a case a transition to the para-electric state takes place in the sample, and the domain pattern is erased. If a sample is externally heated and then allowed to cool, new domains are created on passing the transition temperature. The way in which this happens is of course important for the understanding of a domain pattern that appears in a particular crystal. The creation of domains at the transition temperature is linked to nucleation, which depends on the overall state of the crystal. Most crystals are in a state of non-uniform strain, as a consequence of the conditions during their growth. The domain configuration in a crystal is therefore a result of a compromise between the energy requirements of a perfect crystal and the perturbing effects of strain, defects, and conductivity in the actual crystal. The number and the distribution of vacancies, dislocations, impurities and the doping affect the uniformity of the polarization.

Our experiments with the SEM showed that after cooling the sample to room temperature the domain pattern had the same overall shape and the same polarization directions as before heating. However, if the sample is strained during the cooling the domain pattern can change drastically; see *fig. 7*. The domain pattern that forms during the change of state can of course also be affected by an external electric field. In a particular experiment an electric field was applied parallel to the axis of spontaneous polarization in the sample. This was done with the aid of a gold contact deposited on the sample by vapour deposition. Observations showed that the number of points at which domains started to grow and the velocities of the moving walls both increase with the strength of the applied electric field. So far the discussion of voltage-contrast observation in the dynamic mode has been limited to its use with single-crystal materials. Nevertheless, an extension of the technique to the observation of domains in ferroelectric ceramic materials — which are polycrystalline — is of great practical importance. In the next part of this article results will be discussed



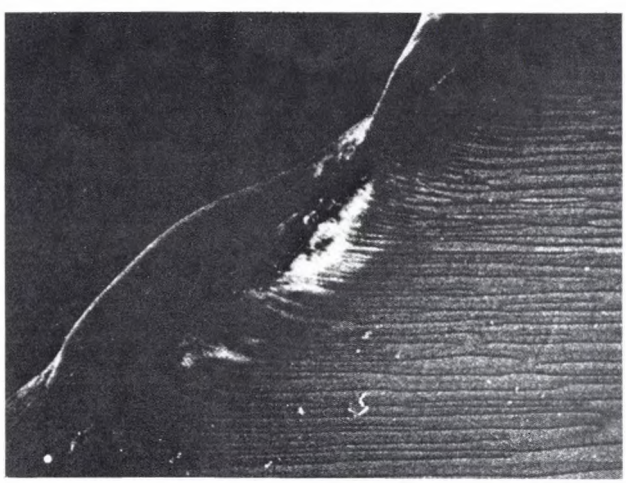


Fig. 7. Photomicrograph of the edge of a strained single crystal of TGS. The sample was in the form of a thin wafer. Two typical domain patterns can be seen.

that were obtained with the SEM in the study of one of the ceramic materials mentioned earlier, lead zirconate-titanate with lanthanum.

Domains in a polycrystalline ferroelectric

Lead zirconate-titanate doped with lanthanum (PLZT) is a ferroelectric ceramic whose crystals are of the perovskite type [5]. This type of crystal can be polarized along several axes: in the unpolarized state, above the transition temperature, these axes are all equivalent. The material that is mostly studied, on account of its unusual electro-optic properties — which make it suitable for shutters, light modulators, image storage and display, as well as for holographic memories [8] — contains a doping concentration of eight atomic per cent of lanthanum to give transparency; the lanthanum is substituted for lead. Zinc and titanium are present in the ratio 65 : 35. This compound possesses a number of physically complicated but generally useful electrical states involving some rather unusual phase transitions [9]. The actual state in a particular sample depends on the thermal and electrical history. The best material is produced by hot pressing and has small grains (1-2 μm). The conventional techniques for the observation of the microstructure have too small a resolution for these grain dimensions. Again, using the SEM in the voltage-contrast dynamic mode proved to be a successful alternative. The samples were polished wafers with a thickness of about 250 μm. The material is hard and suffers no beam damage from a current of 10^{-10} A and accelerating voltages up to 25 kV. However, there was a charging effect, visible on the television monitor, which seriously limited the observation of domain structures. In fact, after a few scans, the electric field induced across the wafer by the excess of surface charge was so high that it generated small intergranular cracks in the ceramic. The various stages of the attack of the surface and the propagation of the cracks across it could all be observed on the monitor as they took place. In one extreme case (a scan of more than 10 seconds) the transverse field due to the surface charge was high enough to eject a crystal grain from its site, leaving a cavity with a well defined preferential crystallographic orientation (*fig. 8*). This effect which has only been observed above the transition temperature, could perhaps be due to the strong electromechanical force that arises on the polarization reversal of each grain. By coating the sample with a conducting layer, the electric transverse field could be eliminated, and degradation of the surface did not then occur. Under these conditions voltage-contrast imaging was

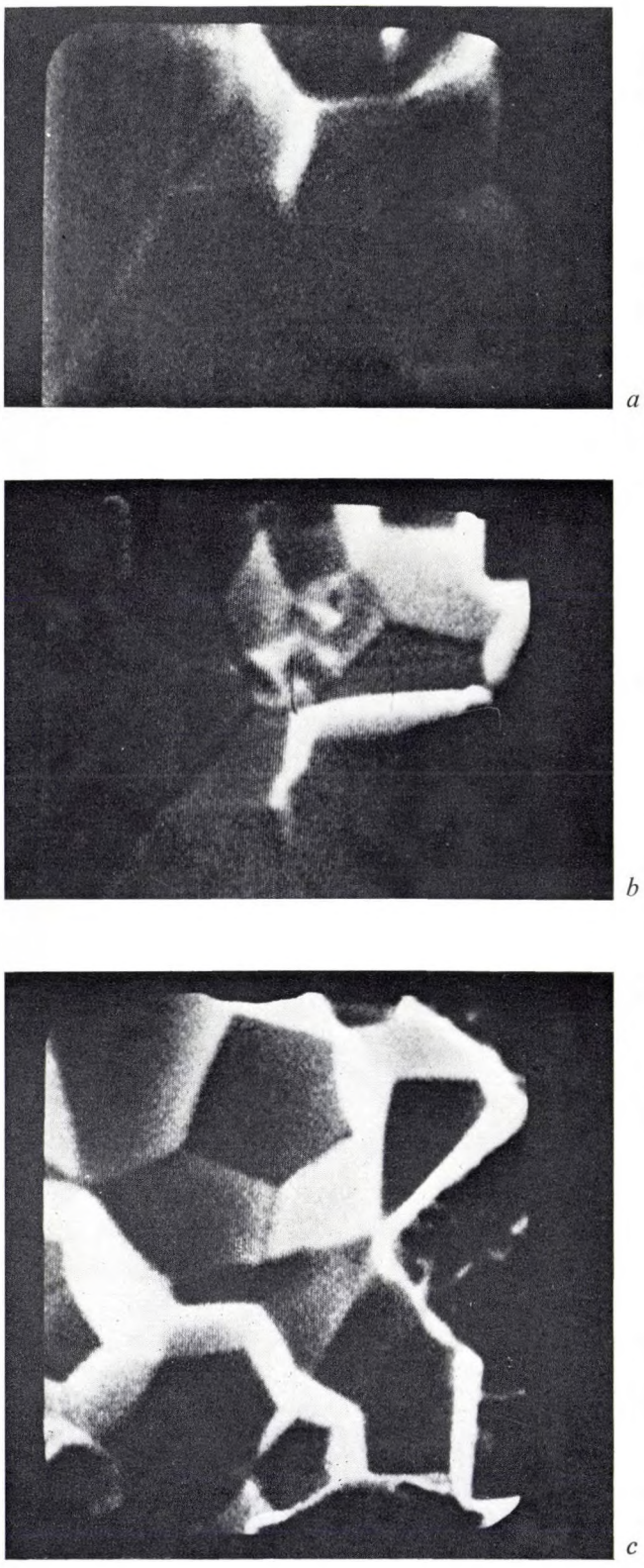


Fig. 8. Beam-induced degradation of the surface of an uncoated sample of PLZT, and the propagation of the cracks. The images were produced with no significant time delay. The scanning time was more than 10 seconds. The small cracks arise between the separate grains of the material and are caused by the high electric field-strengths induced by the surface charges. *a*, *b*) Formation of the cracks. *c*) After a crystal grain has been ejected from its site, a crystallographically well defined cavity remains. If the surface is coated with a conducting layer surface degradation does not occur.

[8] H. N. Roberts, Appl. Optics **11**, 397, 1972.
W. R. Salaneck, J. appl. Phys. **43**, 4468, 1972.

[9] E. T. Keve and A. D. Annis, Ferroelectrics **5**, 77, 1973.



a *b* *c*

Fig. 9. Images of chemically etched lanthanum-doped lead zirconate-titanate (PLZT) of composition 8/65/35 (see text). *a*) Coarse-grained material, electrically poled. *b*) Coarse-grained material, electrically depoled. *c*) Fine-grained material, thermally depoled.

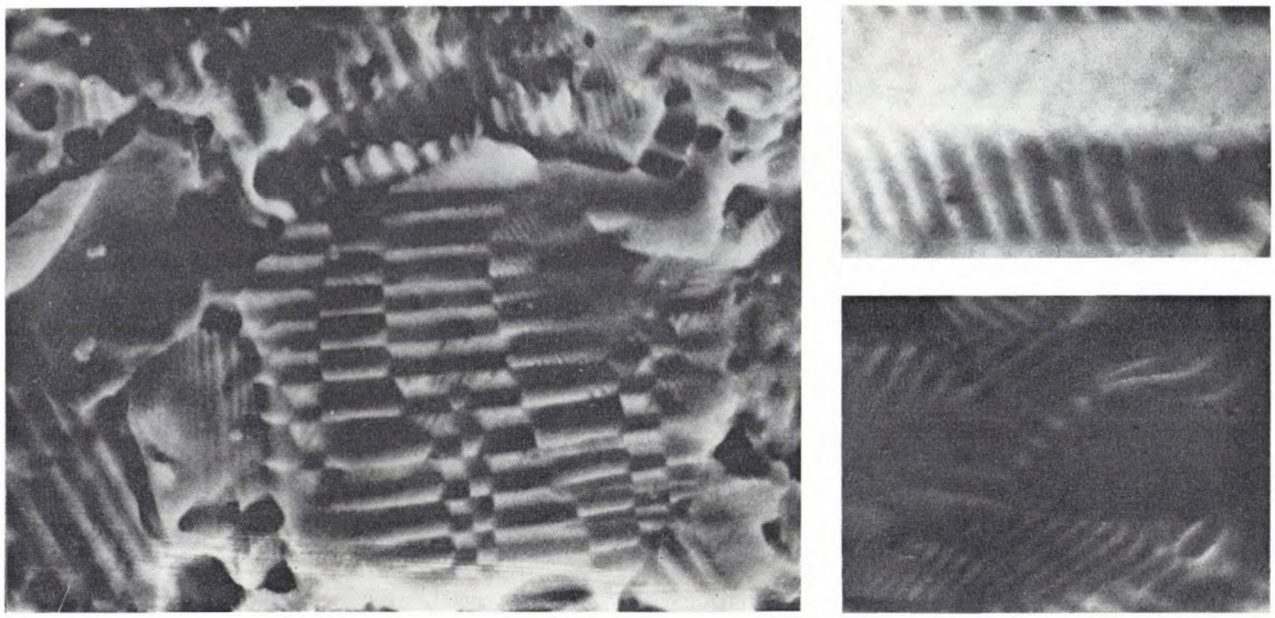
also possible: differences in the electrical and thermal history of the material were found to give clearly observable differences in the microstructure of the grains. The absence of grain structure in thermally depoled material supports E. T. Keve's model of a non-polar cubic structure [9]. For material that was electrically polarized or strained, striations with a spacing generally smaller than 1 μm were observed. Such a result suggests a preferential orientation of the polarization. For electrically depoled material, the images obtained by voltage contrast clearly suggested the presence of randomly distributed domains. The contrast of the different patterns was very poor and found to depend critically on the thickness of the coating. Better results were in fact obtained with chemically etched ceramic wafers with no coating. Surface microstructures of three different states, electrically poled, electrically depoled, and thermally depoled, are shown in *fig. 9*. The grain pattern of a strain-induced phase shown in *fig. 10a* shows complex superimposed striated patterns, with a high degree of preferential orientation that can no longer be satisfactorily explained by considering an ordered ferroelectric domain structure alone. In further investigations with surface replicas of the same sample with a transmission electron microscope, the greater resolution revealed the presence of a twin band structure about 0.01 μm wide (fig. 10*b*). Preliminary investigations suggest an ordered distribution of extended defects. In this case, voltage-contrast imaging in the dynamic mode, combined with information obtained by static etching techniques, might lead to a better understanding of the unusual phase transitions that arise here.

Magnetic-contrast imaging

It seems rather obvious to try to extend the same methods to ferromagnetic materials. In this case imaging is produced by means of magnetic contrast, i.e. by the operation of the Lorentz force, which will deflect the secondary electrons emitted from the sample as they travel through the magnetic fringing fields. In our study of the possibilities that this method might offer we have examined single crystals of magnetoplumbite. This material has uniaxial anisotropy, which implies that the domain structure is rather simple. The saturation magnetization is lower than that of cobalt, but higher than that of the orthoferrites. *Fig. 11* shows the domain structure at the basal surface of a magnetoplumbite sample. The black and white areas reveal the presence of opposing internal demagnetizing fields. The magnetic-contrast response can be maximized by orienting the sample in such a way that the internal domains are directed towards the SEM collector: this indicates that the fringing fields contain a component parallel to the surface and perpendicular to the line between sample and collector [1]. In effect the Lorentz force tilts the distribution of secondary electrons leaving the surface. This tilt occurs about an axis parallel to the component of the fringing field mentioned above. Magnetic-contrast imaging can therefore be improved by increasing the directional sensitivity of the detector [10].

The work on magnetoplumbite has led us to the view that it will be difficult to obtain good contrast, even with optimum orientation conditions. The technique can of course only be used with materials that have a sufficiently high anisotropy to produce fringing fields outside the surface. These various factors set a limitation to the usefulness of the technique, particularly if we compare it with the analogous case of voltage contrast. This was why no further studies of magnetic-contrast imaging were made.

Our experience with this particular case does however enable us to make a few further comments on this method of imaging. We found that the observed



a *b*

Fig. 10. Microstructure of an individual grain of PLZT, of composition 8/65/35 and chemically etched. The crystals have the rhombic structure (β, see also Table I), as a result of strain. *a*) Sets of clearly delineated overlapping striations, as observed by scanning electron microscopy. *b*) Two images of the same striations, but now greatly magnified. These two pictures were observed by transmission electron microscopy, in a conventional electron microscope, of replicas of the surface. The spacing of the striations is less than 0.01 μm. The high degree of preferential orientation cannot be explained from the ferroelectric domain structure alone.

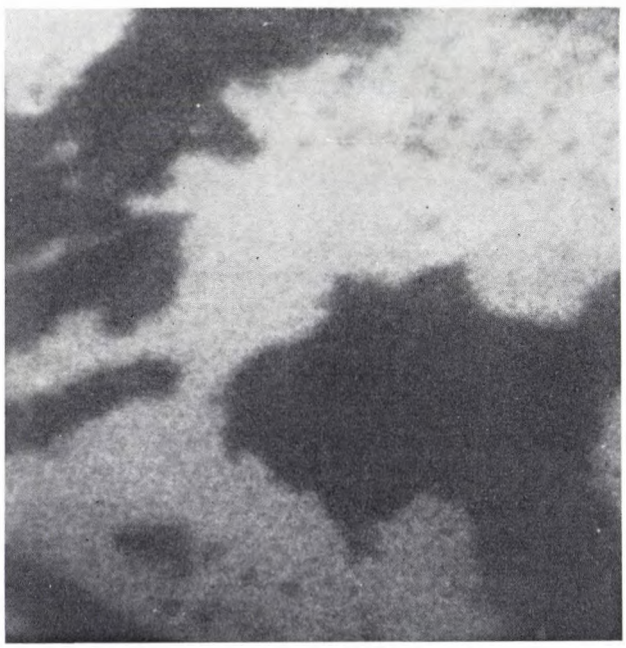


Fig. 11. An example of a magnetic-contrast image made with an SEM. The darker and lighter regions represent ferromagnetic domain structure on the basal surface of a single crystal of magnetoplumbite. Because of Lorentz forces, the fringing fields deflect the secondary electrons emitted from the sample. These deflections produce the contrast in the image.

magnetic-domain pattern is not affected by the surface-charge accumulation during the SEM observation. Since the interaction between the magnetic field and the secondary electrons is direction-dependent, the contrast is related to the orientation of the sample. Finally, magnetic contrast with a good signal-to-noise ratio requires materials with a strong magnetization, such as magnetoplumbite, the example in our study.

An attractive feature of the technique is that no special preparation of the sample is required, as in Lorentz microscopy [11]. The information is obtained almost instantaneously and with a much higher spatial resolution than with the conventional Kerr-effect technique, or the colloidal or X-ray-topography techniques — both of which are unsuitable for dynamic experiments.

Summary. The feasibility of voltage-contrast imaging in scanning electron microscopy is demonstrated by a study of ferroelectric domain configurations. The resolution limit is 0.02 μm. The method is non-destructive. Domain imaging is possible in a static mode, with thinly coated samples, and in a dynamic mode, in which the beam itself induces switching in the domains and the samples are uncoated. Single crystals of TGS (triglycine sulfate), $BaTiO_3$, $Bi_4Ti_3O_{12}$ and the small-grained polycrystalline (ceramic) material 8/65/35 PLZT, which is also ferroelectric, have been studied. Domain patterns with 180° walls and 90° walls are clearly visible. On increasing the beam current (up to 10^{-9} A) local heating beyond the transition temperature can occur and 'writing' with the beam is possible. The analogous magnetic-contrast imaging with ferromagnetic materials seems to be of limited usefulness.

[10] G. A. Wardly, J. appl. Phys. **42**, 376, 1971.
[11] R. Carey and E. D. Isaac, Magnetic domains and techniques for their observation, English Univ. Press, London 1966.

Recent 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*
Mullard Research Laboratories, Redhill, Surrey, England *M*
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*
MBLE Laboratoire de Recherches, 2 avenue Van Becelaere, 1170 Brussels (Boitsfort), Belgium *B*
Philips Laboratories, 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A. (by contract with the North American Philips Corp.) *N*

Reprints of most of these publications will be available in the near future. Requests for reprints should be addressed to the respective laboratories (see the code letter) or to Philips Research Laboratories, Eindhoven, The Netherlands.

W. Albers & L. A. H. van Hoof: Enhanced conversion of x-rays into visible light in thin composite layers (composite anthracene-$PbCl_2$ screens).
J. Electrochem. Soc. **121**, 1341-1345, 1974 (No. 10). *E*

H. Bex: New broadband balun.
Electronics Letters **11**, 47-48, 1975 (No. 2). *A*

J. H. den Boef & J. C. M. Henning: A strain-modulated electron spin resonance spectrometer.
Rev. sci. Instr. **45**, 1199-1202, 1974 (No. 10). *E*

J. Bootsma: The gas liquid interface and the load capacity of helical grooved journal bearings.
Trans. ASME F (J. Lubr. Technol.) **95**, 94-100, 1973 (No. 1). *E*

J. Bootsma: The gas-to-liquid interface of spiral groove journal bearings and its effect on stability.
Trans. ASME F (J. Lubr. Technol.) **96**, 337-345, 1974 (No. 3). *E*

J. Bootsma: Spherical and conical spiral groove bearings: Part I. Theory, Part II. Load capacity and stability.
Trans. ASME F (J. Lubr. Technol.) **97**, 236-242, 243-249, 1975 (No. 2). *E*

J. J. v.d. Broek: Calculation of the formation volumes of alloys.
3rd Int. Conf. on Chemical thermodynamics, Baden near Vienna 1973, pp. 52-58; 1974. *E*

K. H. J. Buschow: Note on the magnetic properties of some Fe_2P-type rare-earth intermetallic compounds.
J. less-common Met. **39**, 185-188, 1975 (No. 1). *E*

C. Crevecoeur & H. J. de Wit: The growth of anodic aluminum oxide layers after a heat-treatment.
J. Electrochem. Soc. **121**, 1465-1474, 1974 (No. 11). *E*

P. A. Devijver: Entropie quadratique et reconnaissance des formes.
Proc. NATO ASI Computer oriented learning processes, Bonas 1974, pp. 39-55. *B*

H. Durand: L'avenir des cellules solaires au silicium à usage terrestre.
Onde électr. **55**, 161-166, 1975 (No. 3). *L*

M. J. C. van Gemert & W. H. de Jeu: On the thin cell method in time domain spectroscopy.
Chem. Phys. Letters **29**, 287-289, 1974 (No. 2). *E*

A. A. van der Giessen: Advances in magnetic recording materials.
Rev. Phys. appl. **9**, 869-876, 1974 (No. 5). *E*

J. Haisma, G. Bartels, W. F. Druyvesteyn, U. Enz, J.-P. Krumme & A. G. H. Verhulst: Observation of various types of bubbles bounded by a compensation wall.
IEEE Trans. **MAG-10**, 630-633, 1974 (No. 3). *E, H*

C. M. Hart, A. Slob & H. E. J. Wulms: Bipolar LSI takes a new direction with integrated injection logic.
Electronics **47**, No. 20, 111-118, Oct. 3, 1974. *E*

G. Jötten, K. Kyser (both with C. H. F. Müller GmbH, Hamburg) **& W. J. Oosterkamp:** X-ray source for mammography.
Medicamundi **19**, 25-27, 1974 (No. 1). *E*

D. Kasperkovitz: Analysis and improvement of a static shift register.
Microelectronics and Reliability **13**, 501-515, 1974 (No. 6). *E*

J. A. Kerr (Mullard Hazel Grove, Ltd., Stockport, Cheshire, England) **& F. Berz:** The effect of emitter doping gradient on f_T in microwave bipolar transistors.
IEEE Trans. **ED-22**, 15-20, 1975 (No. 1). *M*

J. E. Knowles: The origin of the increase in magnetic loss induced by machining ferrites.
IEEE Trans. **MAG-11**, 44-50, 1975 (No. 1). *M*

E. Krätzig & M. Rosenkranz: Ultrasonic determination of the energy gap in superconducting layers.
Phys. Stat. sol. (b) **67**, K 19-21, 1975 (No. 1). *H*

F. E. J. Kruseman Aretz: Doelstellingen en achtergronden van de operating systems PICO, MICRO en MILLI.
Informatie **16**, 672-678, 1974 (No. 12). *E*

H. K. Kuiken: The cooling of a low-heat-resistance sheet moving through a fluid.
Proc. Roy. Soc. London A **341**, 233-252, 1974 (No. 1625). *E*

M. Laguës & J. L. Domange (E.N.S.C.P., Paris): Surface segregation. Comparison between theory and experiment.
Surface Sci. **47**, 77-85, 1975 (No. 1). *L*

F. Meyer & J. J. Vrakking: In-depth information from Auger electron spectroscopy.
Surface Sci. **45**, 409-418, 1974 (No. 2). *E*

F. Meyer & J. J. Vrakking: Comment on 'The adsorption of oxygen on silicon (111) surfaces' by R. Dorn, H. Lüth and H. Ibach.
Surface Sci. **46**, 287-289, 1974 (No. 1). *E*

D. Meyer-Ebrecht & G. Schröder: Rechneranpassung frequenzanaloger Meßsysteme.
Acta IMEKO 1973, pp. 701-712; 1975. *H*

A. Mircea, E. Constant (Université de Lille) **& R. Perrichon** (Univ. de Lille): FM noise of high-efficiency GaAs IMPATT oscillators and amplifiers.
Appl. Phys. Letters **26**, 245-248, 1975 (No. 5). *L*

A. van Oostrom: Influence of adsorbates and electric field on the nucleation and growth of microtips in a vacuum gap.
Proc. 6th Int. Symp. on Discharges and electrical insulation in vacuum, Swansea 1974, pp. 49-70. *E*

K. J. van Oostrum & H. F. Premsela: Visualisation of diffraction information in a TEM image using colour television display.
8th Int. Congress on Electron microscopy, Canberra 1974, Vol. I, pp. 100-101. *E*

I. Pockrand & J. Verweel: Magnetic domains in thin sputtered FeSi films, I. Edge effects and influence of the substrate temperature and the Ar pressure during sputtering.
Phys. Stat. sol. (a) **27**, 413-427, 1975 (No. 2). *H*

H. Rau: Vapour composition and Van der Waals constants of arsenic.
J. chem. Thermodyn. **7**, 27-32, 1975 (No. 1). *A*

D. Rossier & F. Dumont: Dependence of the characteristics of a photoconductor-dielectric optical converter on photogeneration and charge transfer parameters of the photoconductor.
J. appl. Phys. **46**, 849-854, 1975 (No. 2). *L*

E. Roza: Analysis of phase-locked timing extraction circuits for pulse code transmission.
IEEE Trans. **COM-22**, 1236-1249, 1974 (No. 9). *E*

T. E. Rozzi & J. H. C. van Heuven: Quasi-power algebraic invariants of linear networks.
IEEE Trans. **CAS-21**, 722-728, 1974 (No. 6). *E*

G. B. Scott, D. E. Lacklison & J. L. Page: The effects of octahedral Fe^{3+} and tetrahedral Fe^{3+} dilution on the Faraday spectra of bismuth-doped iron garnets.
J. Physics C **8**, 519-529, 1975 (No. 4). *M*

J. L. Sommerdijk, A. Bril & A. W. de Jager: Luminescence of Pr^{3+}-activated fluorides.
J. Luminescence **9**, 288-296, 1974 (No. 4). *E*

A. L. N. Stevels & W. Kühl: New phosphors for X-ray image intensifier tubes.
Medicamundi **19**, 3-7, 1974 (No. 1). *E*

B. J. Stocker: AES and LEED study of the activation of GaAs-Cs-O negative electron affinity surfaces.
Surface Sci. **47**, 501-513, 1975 (No. 2). *M*

T. J. B. Swanenburg & J. Wolter: Frequency dependence of the transmission of high-frequency phonons from a solid into liquid helium.
Phys. Rev. Letters **33**, 882-885, 1974 (No. 15). *E*

T. L. Tansley: AC profiling by Schottky gated cloverleaf.
J. Physics E **8**, 52-54, 1975 (No. 1). *M*

M. J. J. Theunissen: Charge transfer devices: I. Physical principles, II. Applications.
Onde électr. **54**, 317-324, 405-413, 1974 (Nos. 7 & 8). *E*

H. J. Tolle & R. Memming: Fast imaging process in PbI_2 films.
Appl. Phys. Letters **26**, 349-351, 1975 (No. 6). *H*

T. S. te Velde: Mathematical analysis of a heterojunction, applied to the copper sulphide – cadmium sulphide solar cell.
Solid-State Electronics **16**, 1305-1314, 1973 (No. 12). *E*

J. M. P. J. Verstegen (Philips Lighting Division, Eindhoven): A survey of a group of phosphors, based on hexagonal aluminate and gallate host lattices.
J. Electrochem. Soc. **121**, 1623-1627, 1974 (No. 12).

J. M. P. J. Verstegen (Philips Lighting Division, Eindhoven): The luminescence of Tb^{3+} in borates of the composition $X_2Z(BO_3)_2$ ($X = Ba, Sr, Ca$; $Z = Ca, Mg$).
J. Electrochem. Soc. **121**, 1631-1633, 1974 (No. 12).

J. M. P. J. Verstegen, D. Radielović & L. E. Vrenken (Philips Lighting Division, Eindhoven): A new generation of 'deluxe' fluorescent lamps, combining an efficacy of 80 lumens/W or more with a color rendering index of approximately 85.
J. Electrochem. Soc. **121**, 1627-1631, 1974 (No. 12), & **122**, 843, 1975 (No. 6) *(reply to discussion)*.

J. M. P. J. Verstegen (Philips Lighting Division, Eindhoven) **& J. L. Sommerdijk:** Line emission of $SrBe_2Si_2O_7:Eu^{2+}$ and $BaBe_2Si_2O_7:Eu^{2+}$.
J. Luminescence **9**, 297-301, 1974 (No. 4). *E*

J. F. Verwey & R. P. Kramer: Atmos — an electrically reprogrammable read-only memory device.
IEEE Trans. **ED-21**, 631-636, 1974 (No. 10). *E*

J. O. Voorman: Ideal frequency modulator.
Electronics Letters **10**, 387-388, 1974 (No. 18). *E*

P. van der Wurf: On the spectral density of a cyclostationary process.
IEEE Trans. **COM-22**, 1727-1730, 1974 (No. 10). *E*

F. Zernike: Luneburg lens for optical waveguide use.
Optics Comm. **12**, 379-381, 1974 (No. 4). *N*

Contents of Philips Research Reports **30**, Special issue in honour of C. J. Bouwkamp, 1975:

H. B. G. Casimir: Laser modes; an imperfect tribute to C. J. Bouwkamp (pp. 1*-4*).

O. Bottema (Delft): A lesson in elementary geometry (pp. 5*-13*).

C. H. Papas (U.S.A.): On the equation of motion in electrodynamics (pp. 14*-19*).

M. Kac (U.S.A.): An example of 'counting without counting' (pp. 20*-22*).

A. Erdélyi (U.K.): Fourier transforms of integrable generalized functions (pp. 23*-30*).

M. S. Klamkin (Canada): Asymptotic heat conduction in arbitrary bodies (pp. 31*-39*).

J. H. van Lint (T.H. Eindhoven) **& H. O. Pollak** (U.S.A.): An asymmetric contest for properties of arbitrary value (pp. 40*-55*).

I. Kay (U.S.A.): Near and far field HF radar ground wave return from the sea (pp. 56*-64*).

N. G. van Kampen (Utrecht): The collapse of the wave function (pp. 65*-73*).

B. R. A. Nijboer (Utrecht): On a relation between the scattering cross-section in dense media and the energy of a dilute electron gas (pp. 74*-82*).

J. B. Keller (U.S.A.): Effective conductivity, dielectric constant and permeability of a dilute suspension (pp. 83*-90*).

P. Delsarte, J. M. Goethals & J. J. Seidel (T.H. Eindhoven): Bounds for systems of lines, and Jacobi polynomials (pp. 91*-105*). *B*

V. Belevitch & Y. Genin: Reciprocity invariants in equivalent networks (pp. 106*-121*). *B*

J. B. Alblas (T.H. Eindhoven): Relaxation phenomena in electro-magneto-elasticity (pp. 122*-139*).

J. W. Miles (U.S.A.): Asymptotic approximations for oblate spheroidal wave functions (pp. 140*-160*).

J. Boersma (T.H. Eindhoven): Analysis of Weinstein's diffraction function (pp. 161*-170*).

J. A. Geurst: Continuum theory for type-*A* smectic liquid crystals (pp. 171*-186*). *E*

L. B. Felsen (U.S.A.): Complex rays (pp. 187*-195*).

H. Freudenthal (Utrecht): On the cardinality of finite Tits geometries (pp. 196*-204*).

R. L. Brooks (U.K.), **C. A. B. Smith** (U.K.), **A. H. Stone** (U.S.A.) **& W. T. Tutte** (Canada): Leaky electricity and triangulated triangles (pp. 205*-219*).

P. J. Federico (U.S.A.): The number of polyhedra (pp. 220*-231*).

J. Meixner (Aachen) **& S. Sche** (Aachen): Some remarks on the treatment of the diffraction through a circular aperture (pp. 232*-239*).

H. Levine (U.S.A.): Acoustical diffraction radiation (pp. 240*-276*).

K. M. Adams (Delft): The non-amplification property of networks consisting of *n*-terminal resistive devices (pp. 277*-287*).

F. E. J. Kruseman Aretz & J. A. Zonneveld: FFT algorithms (pp. 288*-301*). *E*

A. T. de Hoop (Delft): The *N*-port receiving antenna and its equivalent electrical network (pp. 302*-315*).

H. J. Butterweck (T.H. Eindhoven): Noise voltages of bulk resistors due to random fluctuations of conductivity (pp. 316*-321*).

A. J. Dekkers: *N*-omino enumeration (pp. 322*-328*). *E*

A. J. W. Duijvestijn (Enschede): Fast calculation of inverse matrices occurring in squared-rectangle calculation (pp. 329*-336*).

N. G. de Bruijn (T.H. Eindhoven) **& D. A. Klarner** (U.S.A.): A finite basis theorem for packing boxes with bricks (pp. 337*-343*).

F. L. H. M. Stumpers (Bochum): Some notes on the correspondence between Sir Edward Appleton and Balth. van der Pol (pp. 344*-356*).

N. Marcuvitz (U.S.A.): Eigenmodes, quasimodes and quasiparticles (pp. 357*-375*).

PHILIPS TECHNICAL REVIEW

VOLUME 36, 1976, No. 2

Acoustic surface-wave bandpass filters

D. W. Parker, R. G. Pratt, F. W. Smith and R. Stevens

It is only ten years since the invention of the interdigital transducer made the use of acoustic surface waves a practical possibility in systems for signal processing. Since that time many laboratories have been working on a wide range of circuit elements based on acoustic surface waves. At Mullard Research Laboratories a group of scientists have devised computer methods for designing filters that have a specified frequency response. One of the devices that have been developed with the aid of these methods is an analogue bandpass filter suitable for use as an intermediate-frequency filter in television receivers.

The interdigital transducer, two comb-like electrodes interlocking with one another, allows electrical signals to be efficiently and linearly converted into acoustic signals in piezoelectric material; the acoustic signals propagate in the form of surface waves.

If two such transducers are applied to a piezoelectric substrate, the acoustic surface waves generated by the first transducer can be reconstituted to an electrical signal in the second transducer, giving a filter, as we shall explain in more detail later. The frequency characteristic of the filter depends on the geometry of the two transducers [1].

The designation 'acoustic' is perhaps a little misleading, since it might suggest that the surface waves had frequencies in the audio range (about 20 to 18 000 Hz). In fact acoustic surface waves can have very high frequencies (up to several GHz); the feature they have in common with sound waves is that the signals propagate via oscillations of particles in the material. The velocity of propagation of the surface waves is a few km/s, and the wavelength varies from about 1 mm at 3 MHz to 1 μm at 3 GHz. From the values quoted for the wavelength it can be seen that the dimensions of acoustic surface-wave filters will be of the order of millimetres; they are therefore about 10^5 times as small as their electromagnetic counterparts.

The nature of the acoustic surface-wave filter is such that the initial design costs are relatively high but large numbers of the devices can be produced cheaply. Like integrated circuits — but to a lesser degree — these filters are consequently mainly suitable for mass production.

Our work has been chiefly directed towards the design and manufacture of acoustic surface-wave filters that can be used as i.f. filters in television receivers. Not only are these components that are required in large numbers, but in addition surface waves turn out to be extremely convenient for use in this particular frequency range.

In this article we shall describe how techniques have been developed to design and make such filters and how problems encountered have been solved. Examples of filters we have made, suitable for use in two different television systems, will be given. In addition, we shall describe filters designed for other applications. First of all, however, we shall look more closely at the details of the operation of surface-wave filters and at the relationship between the electrode geometry and the frequency characteristic.

D. W. Parker, B.A., R. G. Pratt, B.Sc., F. W. Smith, B.Sc. and R. Stevens, B.Sc., are with Mullard Research Laboratories, Redhill, Surrey, England.

[1] R. F. Mitchell, Acoustic surface-wave filters, Philips tech. Rev. 32, 179-189, 1971.

Construction and operation of a surface-wave filter

Fig. 1 is a schematic diagram showing the general appearance of a surface-wave filter. The easiest way of understanding its operation is to consider each finger of the two electrodes as a linear source of surface waves, located at the centre-line of that finger. The strength of such an acoustic source in the input transducer depends on the length l by which the fingers overlap; the source strength is to a first approximation directly proportional to this 'finger-length', which can have different values from finger to finger. The acoustic signal that travels over the surface of the substrate therefore consists of components that can be considered as delayed acoustical replicas — of differing strengths — of the electrical input signal, a situation that is characteristic of the 'transversal' filter. At a given place on the surface of the substrate these components will be in phase at certain frequencies and will add to form a strong signal. At other frequencies they will be out of phase and will cancel or almost cancel.

It can be shown that the frequency response $F(\omega)$ of a transducer consisting of a number of equidistant line sources, as in the simple model described above, can be described by a Fourier series whose coefficients are given by the strengths of the acoustic sources. This implies that the frequency response is a periodic function of the frequency. The period Ω of this function is equal to twice the centre frequency ω_0 of the first transmission band; see *fig. 2*. The passbands are identical and they are symmetrical about their centre frequencies. If all the fingers are of the same length and width — i.e. it is a 'uniform' transducer — the first passband is of the form $(\sin x)/x$, where x is given by:

$$x = \frac{(\omega - \omega_0)Np}{2v}.$$

Here N is the number of pairs of fingers and v is the phase velocity of the surface wave. At the centre frequency ω_0 the wavelength of the surface wave is equal to the period p of the interdigital electrodes.

Since in reality the fingers are not true line sources, the real frequency responses differ from the one shown in fig. 2: their heights are markedly different and some are virtually absent. For transducers in which $w = \frac{1}{4}p$, this is for example the case for the passband at $3\omega_0$. Since in practice the higher passbands will usually be situated outside the band passed by the system including the filter, we can confine our attention to the passband around ω_0. The effect of the finger-width on the shape of the first passband is relatively small.

If we assume that that the electrical input signal $e(t)$ is sinusoidal and of the form

$$e(t) = \cos \omega t, \tag{1}$$

then the n^{th} component of the acoustic signal is given by

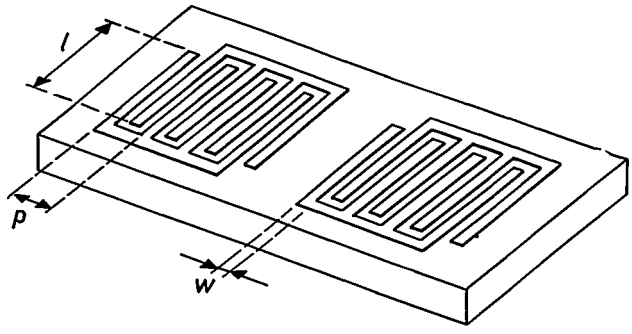


Fig. 1. Simplified representation of a surface-wave filter. On a piezoelectric substrate there are two transducers, each consisting of two comb-shaped electrodes whose fingers enmesh with one another ('interdigital electrodes'). One of the transducers generates surface waves, the other converts them into an electrical signal. In the case shown the distance l, over which the fingers overlap one another, is the same for all fingers ('uniform transducer'). In general this is not the case. In the filters that are discussed in this article the width w of the fingers is usually a quarter of their spacing p.

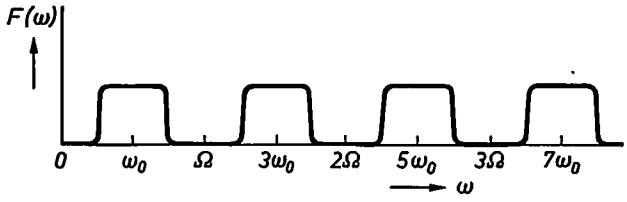


Fig. 2. The frequency response of a hypothetical transducer, consisting of a number of equidistant line sources (spacing $\frac{1}{2}p$), is a periodic function. The period Ω is equal to $2\pi v/p$; v is the propagation velocity of the surface waves. The passbands at ω_0, $3\omega_0$, $5\omega_0$ etc. are all symmetrical in this case and are similar. In a practical transducer, however, the passbands are *not* all similar, because of the effects of the finger width. For example, in transducers for which $w = \frac{1}{4}p$, the band at $3\omega_0$ is missing.

$$a_n(t) = c_n \cos(\omega t + \phi_n), \tag{2}$$

where c_n is the strength of the corresponding source and ϕ_n represents the phase shift introduced when the component $a_n(t)$ is delayed by a time τ_n with respect to $e(t)$:

$$\phi_n = \omega \tau_n. \tag{3}$$

Since we are considering line sources regularly spaced at a distance of $\frac{1}{2}p$, the delay times of the acoustic-signal components are multiples of the time τ in which the waves travel a distance $\frac{1}{2}p$:

$$\tau = p/2v. \tag{4}$$

With the aid of (3) and (4) equation (2) can be expressed as

$$a_n(t) = c_n \cos(\omega t + \omega np/2v). \tag{5}$$

The acoustic signal $s(t)$ is then given by

$$s(t) = \sum_n c_n \cos(\omega t + \omega np/2v). \tag{6}$$

A linear system that responds to an input $e(t) = \cos \omega t$ with an output given by (5) can be described by the frequency response $c_n e^{-j\omega np/2v}$. From this we can conclude that the input transducer, which responds to an input signal $e(t) = \cos \omega t$ with an acoustic signal given by (6), has a frequency response given by

$$F(\omega) = \sum_n c_n e^{-j\omega np/2v}. \tag{7}$$

From (7) it follows that $F(\omega)$ is periodic with a period Ω equal to $4\pi v/p$. The expression (7) can be considered as a Fourier series

expansion of the periodic function $F(\omega)$. For a certain specified frequency response, the strengths c_n of the acoustic sources can be found from the equation

$$c_n = \frac{1}{\Omega}\int_{-\Omega/2}^{\Omega/2} F(\omega)e^{j\omega np/2v}. \quad (8)$$

Since the nature of the surface-wave filter requires the sum of all the source strengths to be zero,

$$\sum_n c_n = 0,$$

d.c. transmission is not possible, and this excludes the lowpass filter from the applications of surface-wave filters.

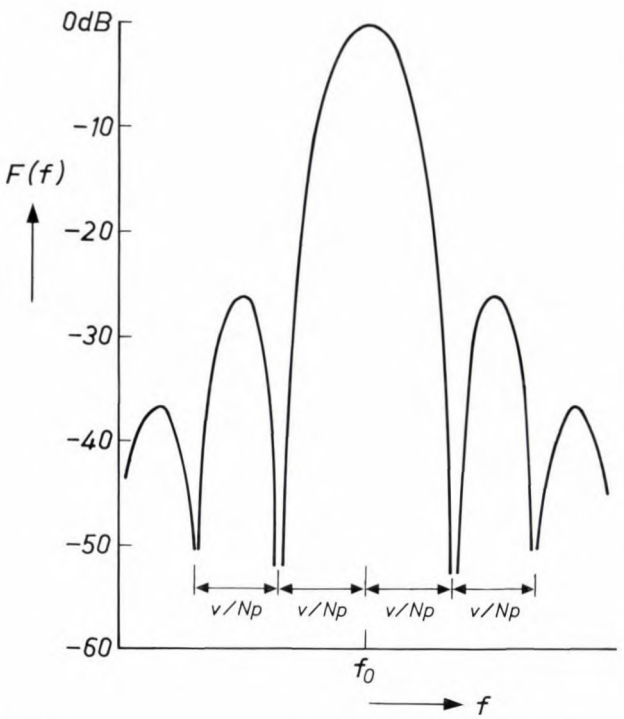


Fig. 3. Amplitude response of a surface-wave filter with uniform transducers. In this case the response is the product of the responses of the input and output transducers. Since both have the form $(\sin x)/x$, the amplitude response has a shape that can be described by the function $\{(\sin x)/x\}^2$. The bandwidth of the filter, measured between the first two zero points from the centre, is $2v/Np$, and the rest of the zeros (points with infinite attenuation) have a spacing of v/Np; N is the number of pairs of fingers.

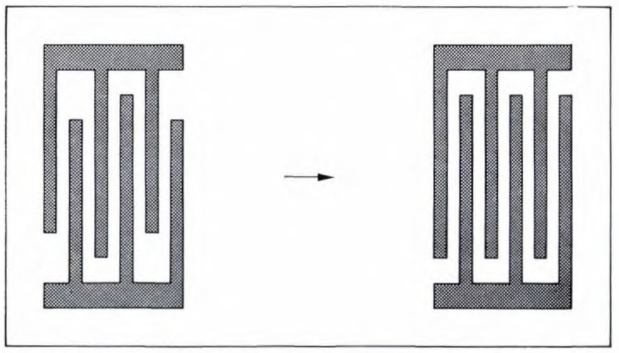


Fig. 4. Surface-wave filter with a single uniform transducer (the receiver) and another one with fingers of dissimilar length. In this case also the frequency response of the filter is equal to the product of the responses of the transducers.

Similar reasoning applies for the receiving transducer. The sensitivity of the pairs of fingers functioning as receivers — like the strength of the acoustic sources — is approximately proportional to the length l. In the case of a uniform receiving transducer with the same number of fingers as the transmitting transducer, the overall filter response is of the form $\{(\sin x)/x\}^2$. The bandwidth of this filter, measured between the two central nulls, will be $2v/Np$; the other nulls will be separated by v/Np, as shown in *fig. 3*.

In the general case in which neither transducer is uniform, the frequency response of the filter is not the same as the product of the frequency responses of the two individual transducers. This is the case, however, if only one of the transducers is non-uniform. It is therefore in general desirable to try to obtain a particular filter response with one uniform and one non-uniform transducer.

Of course, the simple model that we have used here to explain the operation of the filter has to be corrected for such things as diffraction of the acoustic waves and the fact that the sources are not independent of each other. Nevertheless, it provides a means of designing these filters in a systematic way. All that needs to be done to obtain a particular frequency response, in principle, is to find the Fourier series of a periodic repetition of that response. This gives a list of the strengths of the sources required and of their positions. In the following section we shall explain how these procedures are put into practice. We shall also explain how an asymmetrical characteristic can be obtained.

Designing surface-wave filters

Our technique for surface-wave filter design uses a set of computer programs that require very little intervention by the user. A specification of relative amplitude and group delay as a function of frequency (with tolerances) is given, and the computer programs design the filter as far as the point at which layout data can be produced for a standard automatic mask-making system.

The principal method that we have used for variation of the source strength, and hence to obtain the desired frequency response, is to vary the length l by which the fingers overlap; the width w is kept constant. In all the filters we have made, we varied the length of the fingers in only one of the transducers (*fig. 4*): in that case the total frequency response of the filter is the product of the individual responses.

The number of fingers and the centre frequency of the uniform transducer are first chosen by the designer. This is not difficult, since its response is known to be of

the form $(\sin x)/x$. The non-uniform transducer is then designed by the computer such that the overall response lies within the given specifications. In the synthesis program a very simple model of the surface-wave filter is used; each finger of the non-uniform transducer is considered as a source of acoustic waves at a certain strength and position on the substrate. This model is usually adequate for most filters of reasonably wide 'aperture' (i.e. the length of the longest finger of the electrodes) and not too great a bandwidth. Other programs are also available for checking whether second-order effects, such as diffraction, reflection, interaction between sources, etc., will significantly alter the response of the filter.

The first step in the design is to calculate a target response for the non-uniform transducer, together with the permitted tolerances. When this response is repeated periodically, a Fourier series can be found that describes the response sufficiently accurately. This series then indicates the acoustic source strengths required from sources spaced regularly at half a wavelength ($\lambda_0/2$) at the centre frequency ω_0.

Two problems may now arise. In the first place the Fourier series gives in general an infinite number of terms, and hence sources. Secondly, if an amplitude response is required that is asymmetrical about ω_0, the Fourier coefficient will be complex.

In a practical device the number of sources will not of course be infinitely large. This can perturb the response in such a way that it may depart from the specification. Analytical methods exist that allow a specified response to be approximated as well as possible with a limited number of sources. However, the 'best fit' response thus obtained is not always satisfactory, since departure from the specification may be less acceptable at some frequencies than at others. For example, in a television i.f. filter the absolute accuracy required in the trap regions is considerably greater than that in the centre of the passband, where some ripple ($\pm \frac{1}{2}$ dB) is tolerable. Our approach is to limit the number of sources to a value that we have found from experience to give a good chance of meeting the specification. The source strengths are then corrected until the response lies within the given tolerance. In our approach the response of the transducer with the limited number of fingers is compared with the specification and a number of the most serious errors are identified. These are then corrected by 'adding' to the original transducer a number of correcting transducers, one for each error point [2]. Usually, new errors will then appear, so the process is repeated. In practice, it is found either that the response is brought fairly rapidly within the specification, or that the number of points with large errors rapidly increases. The second case indicates that more sources are required and that the design procedure should be started again.

The second problem arises if an asymmetrical amplitude response is required. The fact that complex Fourier coefficients are found in this case means that the sources must generate acoustic signals not only with different amplitudes, but also with different phases. This is not really practical, of course, but an approximation that is valid at one frequency is to give each source a different delay by moving their positions away from the regular $\lambda_0/2$ spacings, thus changing the relative phases. In the design procedure both the strength and position of the sources now have to be varied to bring the filter response within the specification. We have a technique for doing this, but because of the approximations good results are only obtained if the asymmetry of the amplitude response is not too great.

A way out of this problem is obtained by considering the desired asymmetrical response as the left-hand half of a symmetrical one, in the manner indicated in *fig. 5*. The Fourier coefficients of this new function are real, and only the problem of the finite number of sources remains [3]. The transducers produced by this technique have sources spaced at a constant distance of $\lambda_0/4$, and are called double-finger transducers [4]. A disadvantage of the smaller electrode spacing is that the resolution of the fabrication process must be twice as large.

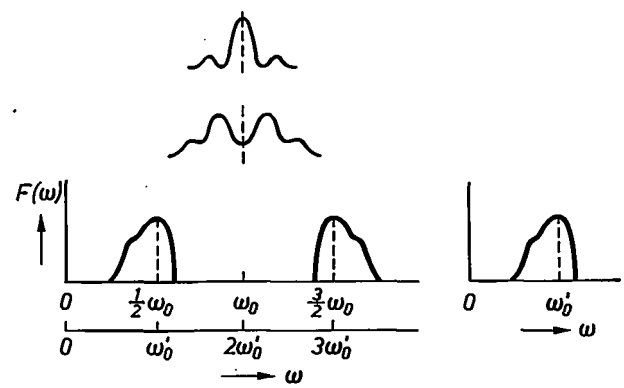


Fig. 5. Illustrating the method for making filters with an asymmetrical frequency response without making use of sources with a phase difference between them (i.e. with a complex source strength). If the distance p (see fig. 1), starting at half the wavelength corresponding to ω_0, is decreased, the central maximum of the first passband (see figs. 2 and 3) becomes lower and eventually the band splits into two parts. Because of the desired symmetry about ω_0 these two parts of the band are mirror images of one another, but each one is itself asymmetrical. If p is chosen such that the partial bands are located at $\frac{1}{2}\omega_0$ and $\frac{3}{2}\omega_0$ and the ratio of w to p in the other transducer is made such that the third harmonic is suppressed, then the filter obtained has an asymmetrical amplitude response about the frequency ω_0' ($= \frac{1}{2}\omega_0$). The finger spacing in the first transducer is then half that in a filter with a symmetrical response about ω_0'.

The phase response

So far we have not considered the phase response of the filter. When the acoustic sources in the transducer are symmetrically arranged about the centre-line of the transducer, the phase response will be linear, irrespective of the amplitude response. The design procedures described above all impose this restriction and therefore only produce filters with linear phase response. Sometimes, however, a prescribed nonlinear phase response is required. Such filters are chiefly used in some European countries as i.f. filters in television receivers, as we shall explain more fully later. We will now give some attention to the way in which such filters are designed.

Our method for designing a filter with a nonlinear phase response is based on the two possible ways that exist for designing a transducer with a linear phase response; see *fig. 6*. The symmetrical transducer (fig. 6*a*) has sources of equal magnitude and sign on either side of the centre-line. Using the Fourier series expansion to describe the frequency response of the symmetrical transducer, we find:

$$F_s(\omega) = B_3 e^{j\omega d_3/v} + B_2 e^{j\omega d_2/v} + B_1 e^{j\omega d_1/v} + B_1 e^{-j\omega d_1/v} + B_2 e^{-j\omega d_3/v} + B_3 e^{-j\omega d_2/v}$$
$$= 2B_1 \cos(\omega d_1/v) + 2B_2 \cos(\omega d_2/v) + 2B_3 \cos(\omega d_3/v),$$

where the strengths of the sources are indicated by $B_{1,2,3}$ and the distances from the centre-line to the source by $d_{1,2,3}$. For the frequency response of the asymmetrical transducer (fig. 6*b*) we find:

$$F_a(\omega) = A_3 e^{j\omega d_3/v} + A_2 e^{j\omega d_2/v} + A_1 e^{j\omega d_1/v} - A_1 e^{-j\omega d_1/v} - A_2 e^{-j\omega d_2/v} - A_3 e^{-j\omega d_3/v}$$
$$= 2jA_1 \sin(\omega d_1/v) + 2jA_2 \sin(\omega d_2/v) + 2jA_3 \sin \omega d_3/v),$$

where the strengths of the sources are indicated by $A_{1,2,3}$ and the polarity by a plus or a minus sign. From both expressions we conclude that the symmetrical transducer has a real frequency response, representing a zero phase shift for all frequencies, while the asymmetrical transducer has an imaginary frequency response (note the factor j), representing a 90° phase shift at all frequencies.

The method used for designing a nonlinear phase filter is as follows. The required phase and amplitude responses are resolved into two components, each with a linear phase response but with a phase difference of 90° with respect to each other. Each of the components can be realized with a single transducer: one with a symmetrical transducer, the other with an asymmetrical transducer. If the source positions are arranged to be the same in both transducers, they can be combined to produce a single transducer, which will have a nonlinear phase response. The combination is done by adding the strengths of corresponding sources, as shown in fig. 6*c*.

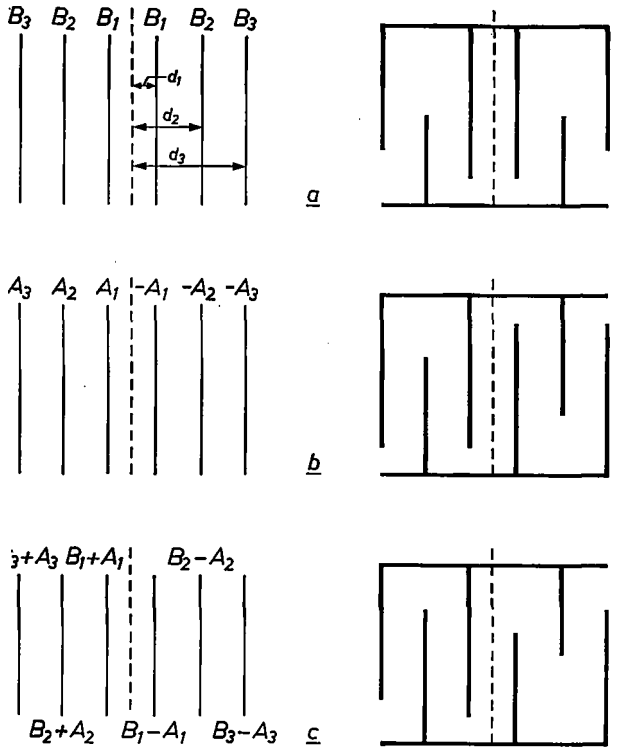


Fig. 6. The diagrams on the right show schematic representations of a transducer with even symmetry (*a*), a transducer with odd symmetry (*b*) and an asymmetrical transducer, formed by combining an odd and an even transducer (*c*). The source strengths are indicated by $A_{1,2,3}$ and $B_{1,2,3}$, the positions of the fingers by $d_{1,2,3}$.

Calculation of the electrode configuration

We have now described how the source strengths and positions are calculated from the specification of the filter response for both linear and nonlinear phase responses. The remaining problem is that of translating this information into the physical dimensions of the actual device. The relative strength of a finger functioning as an acoustic source depends on the electric charge on the finger. This in turn depends on the capacitance between that finger and the fingers of the other electrode.

A good approximation is obtained by assuming that the capacitance is mainly determined by the length by which a finger overlaps the two adjacent fingers. Allowance does however have to be made for some end effects. For example, at the end of a finger a small additional amount of charge is present due to fringing

[2] K. W. Moulding and D. W. Parker, IEEE 1974 Ultrasonics Symp. Proc., p. 168.
[3] R. F. Mitchell and D. W. Parker, Electronics Lett. **10**, 512, 1974.
[4] T. W. Bristol, W. R. Jones, P. B. Snow and W. R. Smith, IEEE 1972 Ultrasonics Symp. Proc., p. 343.

fields to neighbouring electrodes. In addition, the capacitance per unit length at the end of the interdigital array is not the same as at the centre, at least for a long array. All these effects are allowed for in a program that calculates the length of each finger from a list of the required sources.

A particular set of source strengths can only be obtained with a single configuration, and it may be that this structure is 'skewed', as shown in *fig. 7*. This will occur, for example, when the fingers of one of the two electrodes have a relatively short overlapping length with the adjacent finger to the left and a relatively long overlapping length with the finger to the right. A skewed transducer has an unnecessarily large aperture, and hence an increased insertion loss. Such an undesirable configuration can be avoided by synthesizing a set of sources that not only meets the specified frequency response, but also satisfies some special rules for preventing skewing. This method has been found to work in practice and is included in our filter-synthesis program.

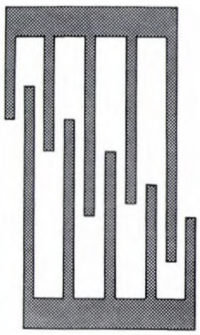


Fig. 7. In seeking an electrode design that will give a particular response, a 'skewed' arrangement may be indicated. This undesired result can be avoided if constraints on the strength of the acoustic sources are included in the computer program.

Filter manufacture

The end result of the design process described in the preceding section is a punched tape suitable for the control of an automatic cutting machine. This produces a mask (the artwork) for the photolithographic process. From there on the electrodes are formed in a sputtered or evaporated metal layer (usually gold or aluminium) on the surface of the substrate by standard processes well known to manufacturers of integrated circuits. Since only one mask is required there are no problems of alignment [*].

However, other problems occur, which are peculiar to surface-wave devices. The electrode systems are large by integrated-circuit standards (of the order of millimetres), and the fingers can be long and thin. Fabrication problems impose a lower limit on the spacing of the fingers, which determines the highest frequency the filter can handle. For conventional ultraviolet photolithography of double-finger transducers on a $Bi_{12}SiO_{20}$ substrate, for example, this highest frequency is about 250 MHz. If an electron beam is used instead of light, this limit can be raised to about 1 GHz. Since the transducers are so large, it is important to ensure that the optical system does not distort the finger shapes as this would degrade the performance of the filter.

The propagation surface of the substrate must be carefully polished to remove scratches and imperfections introduced when it was cut. There are two reasons for this. Firstly, surface scratches may cause breaks in individual fingers, which would affect the frequency response. Secondly, since the energy of the surface wave is concentrated within a layer one or two wavelengths deep, its propagation may be affected by damage to that layer. As an example, the surface wavelength at a frequency of 37 MHz on a (100) plane of $Bi_{12}SiO_{20}$ with propagation in the $\langle 110 \rangle$ direction is only 45 μm. At 370 MHz it is about 4.5 μm, and so on.

Once the filter has been made, it must be packaged, to protect the substrate from dirt and moisture, which would cause damping of the surface waves. The packaging must not of course interfere with the propagation surface.

Choice of substrate material

The performance of the filter depends not only on the geometry of the electrodes but also on the properties of the substrate material. Its size, insertion loss, stability, reproducibility and cost all depend directly on the substrate material. Not all of these factors can be optimized simultaneously and the choice of materials is therefore always a compromise. We shall now look more closely at the more important material properties.

When an electrical signal is applied to a pair of electrodes on a substrate of piezoelectric material, not all of the energy is converted into acoustic waves. Some of the energy is stored in the capacitor formed by the two electrodes. The ratio of the stored to the radiated energy is determined by the square of the electromechanical coupling constant k. The insertion loss is directly dependent on the value of this constant, which means that a material with a high value of k is desirable. (The electrode capacitance can always be tuned out with an inductance, of course, but this is not always convenient and in any case will reduce the bandwidth of the filter.) In single-crystal materials — which are usually used — k is a function of the direction and plane of propagation, given by [5]

$$k^2 = \frac{2(v - v_c)}{v},$$

where v is the free surface velocity and v_c is the velocity when the surface is covered by a massless perfect conductor. Both v and v_c can readily be computed from

the basic constants of the material as functions of the plane and direction of propagation. We have used the approach to calculate, for example, the variation of k in the principal planes of $Bi_{12}GeO_{20}$ and $Bi_{12}SiO_{20}$ [6] [7].

Another important property of the material is the velocity v of the surface wave. The importance of this can best be understood if we consider a filter consisting of two uniform transducers like those shown in fig. 1; all the fingers have the same length l, width w and period p. As mentioned above, the period will be equal to the wavelength at the frequency ω_0 for which the insertion loss is a minimum. In the frequency and temperature range of interest, the velocity does *not* depend on the frequency. The positions of the maxima and the nulls (see fig. 3) will thus depend directly on the velocity, which is fortunately highly reproducible from crystal to crystal (to better than 0.1 %) and does not vary with time, at least in single crystals. The velocity does however vary with temperature. (For $Bi_{12}SiO_{20}$ this, together with the temperature dependence of the dimensions, gives a temperature coefficient for the delay of about $+ 118 \times 10^{-6}/°C$.) It is therefore necessary to choose a material in which the temperature coefficient is low enough for the particular application. If a very high stability is required, it is possible to make use of certain planes in quartz (e.g. the ST cut), for which the first-order term of the temperature coefficient is zero; unfortunately the value of k^2 is then very low. This is all too common a situation, in which a compromise choice has to be made [**].

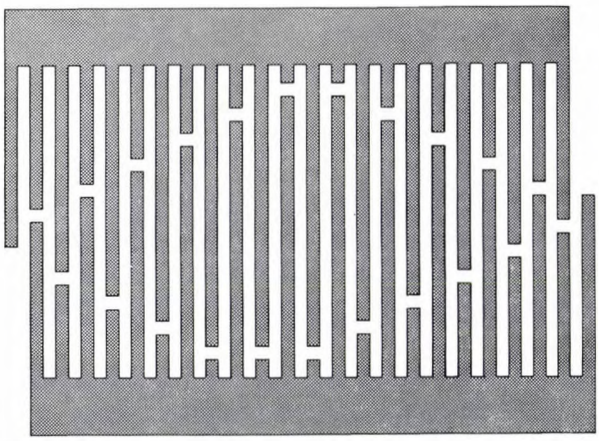


Fig. 8. Transducer with fingers of dissimilar length, which has 'dummy' fingers. Besides the normal active fingers inactive ones are included as well; if these fingers were not present the wavefront of the transmitted wave would not be truly planar.

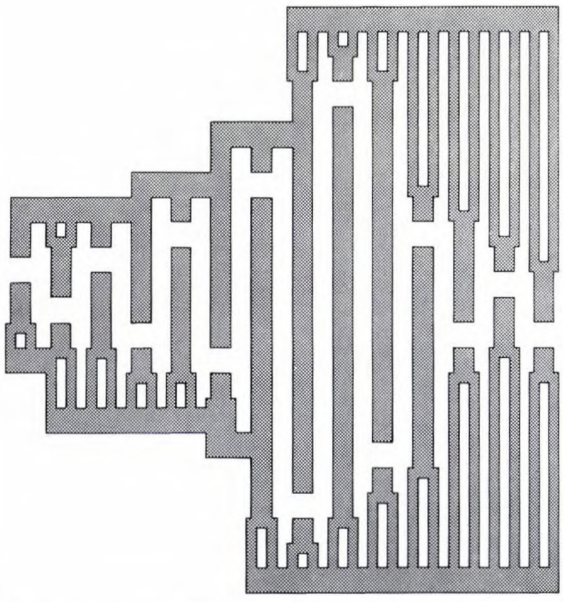


Fig. 9. Transducer with very low interaction between the acoustic sources. In this transducer the inactive fingers are doubled over part of their length, at a spacing of $\frac{1}{4}\lambda_0$, so that the reflected waves cancel.

Problems

In the model used for calculating the filter response it is assumed that each finger launches a plane wave. If the finger is short this assumption is not justified, because of diffraction effects, and corrections are necessary. However, even if the individual fingers do all launch plane waves, the beam that emerges from the transducer will not necessarily have a plane wavefront. This is related to the fact that in a transducer with unequal fingers the waves that propagate in different parts of the aperture do not always pass under the same amount of electrode metal. Since the velocity under the metal layer is less than that on the free surface (by about 1 to 2 %) the wavefront will be distorted. This can be corrected by including inactive or 'dummy' fingers [8] in the transducer, as can be seen in *fig. 8*.

Attention must also be paid to the fact that the waves are partially reflected from the fingers themselves, as a result of the change in velocity at the interface between the free surface and the part covered by the metal. Also the fingers in a transducer can detect and reradiate waves launched earlier, and this can degrade the filter performance, especially the pulse response. This effect can be minimized by using the double-finger configurations. In these designs the finger spacing is $\frac{1}{4}\lambda_0$, so that at the centre frequency the reflections from adjacent fingers are of opposite phase. If, because of fab-

[5] J. J. Campbell and W. R. Jones, IEEE Trans. **SU-15**, 209, 1968.
[6] R. G. Pratt, G. Simpson and W. A. Crossley, Electronics Lett. **8**, 127, 1972.
[7] R. G. Pratt, G. Simpson and C. W. O'Connor, Electronics Lett. **10**, 369, 1974.
[8] R. H. Tancrell and R. C. Williamson, Appl. Phys. Lett. **19**, 456, 1971.
[*] Important contributions to the method of applying the electrodes were made by A. L. Karr of these laboratories.
[**] Lithium iodate ($LiIO_3$) has been reported (V. B. Jipson, J. F. Vetelino, A. Jhunjhunwala and J. C. Field, Proc. IEEE **64**, 568, 1976) as having directions in which the delay temperature coefficient is low but k^2 is high. Unfortunately, it dissolves in water and is thus not suitable for practical filters.

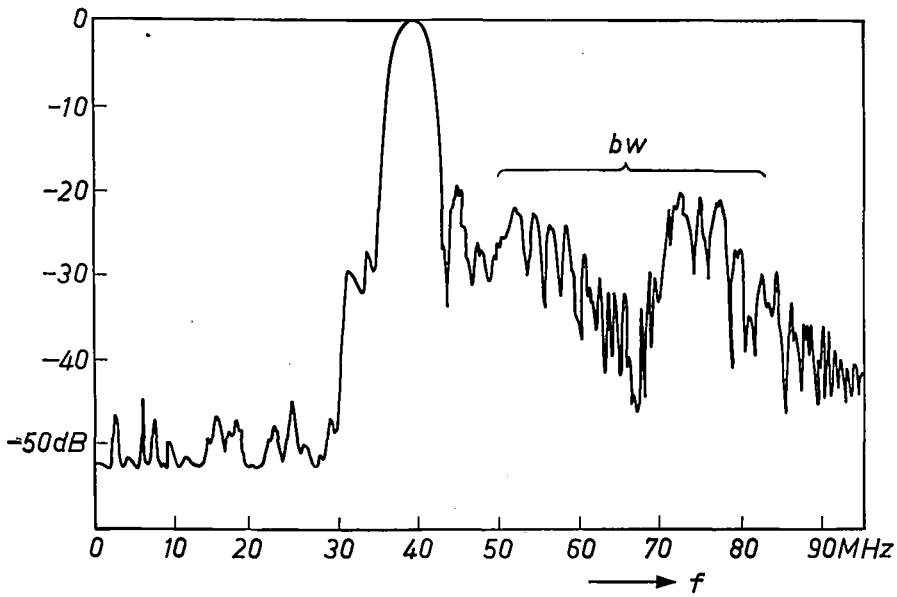


Fig 10. Besides surface waves the transducer can also generate waves that propagate in the interior of the substrate. If these bulk waves can reach the output transducer, e.g. by a reflection from the base of the substrate, they will affect the frequency response of the filter. In the response shown this is the case in the part indicated by *bw*.

rication problems, double-finger designs cannot be used, it is possible to use a compromise design in which the active fingers are single but the inactive ones are double [9]; see *fig. 9*. This considerably reduces the interaction.

The electrode structure used to launch the surface waves can also launch waves into the bulk of the material. These waves can cause unwanted responses at frequencies that depend on their velocity. An example of the effect of these waves on the amplitude response of the filter is shown in *fig. 10*. For a periodic radiating structure waves are launched at an angle θ to the surface, where θ satisfies the usual phase-matching condition:

$$p \cos \theta = n\lambda_B.$$

Here n is an integer and λ_B is the wavelength of the bulk wave. (Strictly, this equality is only valid for an infinite transducer; in a finite transducer the waves are launched over a range of values of θ.) The contribution from the waves that strike the lower surface of the substrate and are reflected from it can be reduced by roughening this surface so that they are scattered. The waves, however, that are launched at small values of θ, so that they reach the receiving transducer without being reflected from the lower surface, cannot be reduced in this way. The only solution is to choose a direction of propagation in which these waves are not too strongly excited [10].

Finally, we have to deal with the fact that in addition to the direct acoustic path from the transmitting to the receiving transducer, there are at least three other ways for an input signal to reach the receiving transducer, each with its own delay time.

The first way in which the input signal can reach the output transducer is direct electrical breakthrough due to the capacitive coupling between the two transducers. This can be reduced by putting an earthed metallic strip on the surface between the two transducers.

The second arises from the fact that the transducers are bidirectional; in fig. 1, for example, it can be seen that the input transducer will launch just as much surface-wave energy to the left as to the right. The wave going to the left will travel to the left-hand end of the substrate and will then be reflected to the output transducer. In a similar way, acoustic waves that pass under the output transducer will be reflected from the right-hand end of the substrate. Both of these waves will arrive at the output transducer with a delay different from that of the wanted signal. These end-reflected signals can usually be reduced to an acceptable level by setting the ends of the substrate at an angle and by putting absorbent material in their path.

The third way in which a delayed signal can arise is through reflection of a signal at the output transducer so that it travels back to the input transducer and is there reflected again back to the output transducer. This doubly reflected 'triple-transit' signal will have three times the delay of the wanted signal. Many methods have been devised to reduce this signal [11], but they all require a larger and more expensive substrate or an increased insertion loss, or both. The level of this triple-transit signal relative to the wanted signal depends on the k value of the substrate material: the larger k^2, the higher the level. It follows that the level of the triple-transit signal increases as the insertion loss of the filter decreases.

Acoustic surface-wave filters as circuit components

Apart from the frequency response, the other features of a surface-wave filter that will be of interest to a user who wishes to include it in an electrical circuit are the impedance and the insertion loss. These factors depend on the substrate material, the bandwidth, the aperture and the nature of the circuit, as will now be explained. We shall assume here that we are dealing with a filter with simple uniform transducers. In the case of non-uniform finger-lengths some modifications to the expressions will be necessary.

At the frequency of its peak response, a transducer can be represented by the simple equivalent circuit shown in *fig. 11a*. The power dissipated in the resistance R is a measure of the power converted into acoustic energy. It can be shown [12] that for a uniform transducer

$$R = \frac{\pi}{4N\omega_0 Ck^2}.$$

Thus for a transducer with a given bandwidth and substrate material, the ratio Q of the resistive part R to the reactive part X of the impedance is given by

$$Q = R/|X| = \omega_0 CR = \pi/4Nk^2.$$

This relation provides a method for finding k^2 by measuring the impedance of the transducer.

The capacitance of the transducer is given by [13]

$$C = \varepsilon_0 \varepsilon_r LN,$$

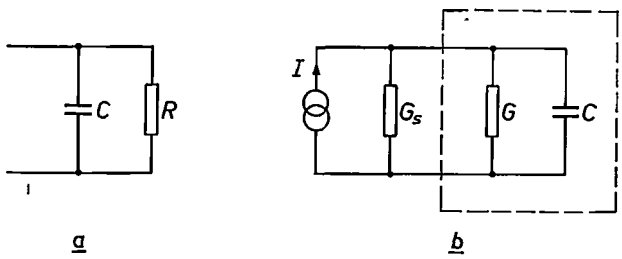
where ε_0 is the permittivity of free space, ε_r is the effective relative permittivity of the substrate material for the propagation plane and direction used and L is the aperture of the transducer. The value of L can be varied within certain limits to give a convenient value for the filter impedance. The lower limit is set by diffraction problems and the upper limit by the acceptable size of substrate.

The insertion loss of the device is made up of two kinds of losses: those not related to matching conditions and those that are. The first category includes losses in the substrate, ohmic losses in the metallization and losses that arise in the propagation of acoustic waves. These can usually be neglected at frequencies below a few hundred MHz for single-crystal materials at room temperature. This first category also includes the losses that arise because the transducers emit acoustic waves in two directions. In a filter with two transducers this bidirectional loss is 6 dB, because each of the interdigital transducers propagates surface waves in two directions with equal power whereas in our filter we only make use of one of these directions.

If the filter is terminated by a resistance there will be a mismatch, since not all the available power from the source can be transferred to the resistive part of the transducer impedance. This introduces a loss of the second category. If both transducers of the filter have the same Q, the minimum value of the mismatch loss for the filter is given by

$$20 \log_{10}\left\{\frac{2}{1+\sqrt{1+Q^2}}\right\} \text{dB}.$$

The magnitude of the losses that will arise with other types of termination can also be calculated easily.



a b

Fig. 11. *a*) Equivalent circuit for a transducer at its centre frequency f_0. C capacitance, R resistance. *b*) Circuit that can be used to determine the fraction of the available power from the signal source transmitted as acoustic power. The source, with conductance G_s, is shown to the left of the dashed line, and the transducer to the right of it ($G = 1/R$).

The expression just given for the minimum value of the loss with a resistive termination can be derived as follows. We consider the simplified circuit shown in fig. 11*b*. The interdigital transducer is represented by its equivalent circuit of a capacitance C in parallel with a conductance G. Power is supplied by a constant-current source I of internal conductance G_s.

The power converted into surface waves is represented by the power dissipated in G. At the frequency ω_0 this is given by

$$P = \frac{I^2 G}{(G_s + G)^2 + \omega_0^2 C^2}.$$

With $\omega_0 C = QG$ this can be rewritten as:

$$P = \frac{I^2 G}{(G_s + G)^2 + Q^2 G^2}.$$

The available power from the source P_A is given by

$$P_A = \frac{I^2}{4G_s},$$

so that the fraction of the available power that is transferred to G is described by

$$\frac{P}{P_A} = \frac{4GG_s}{(G + G_s)^2 + Q^2 G^2}.$$

For maximum power transfer it is required that

$$G_s = \sqrt{G^2 + \omega_0^2 C^2},$$

or

$$G_s = G\sqrt{1 + Q^2}.$$

It follows that

$$\left(\frac{P}{P_A}\right)_{max} = \frac{2}{1 + \sqrt{1 + Q^2}}.$$

[9] D. E. Penna, D. W. Parker, F. W. Smith and R. Stevens, Electronics Lett. **10**, 489, 1974.
[10] R. F. Mitchell and E. Read, IEEE Trans. **SU-22**, 264, 1975.
[11] M. F. Lewis, Electronics Lett. **8**, 553, 1972.
[12] W. R. Smith, H. M. Gerard, J. H. Collins, T. M. Reeder and H. J. Shaw, IEEE Trans. **MTT-17**, 856, 1969.
[13] H. Engan, IEEE Trans. **ED-16**, 1014, 1969.

This is the maximum fraction of the available power that can be transferred to G and hence is the minimum mismatch loss for a single transducer. For a filter the minimum mismatch loss is the square of this, if it is assumed that both transducers have the same value for Q.

Expressing this in decibels, we obtain the above expression.

Surface-wave filter for a television receiver

A block diagram of a television tuner and i.f. section is shown in *fig. 12*. The incoming UHF or VHF signals are converted in the tuner T into signals of frequency about 37 MHz, the i.f. frequency. The tuner itself is not very selective and it is the primary function of the bandpass filter F to separate the wanted signal from any others that may be present. Since one of the sidebands of the television signal is partially suppressed (vestigial-sideband modulation), the shape of the response curve is very important for good demodulation. To prevent intermodulation distortion in the receiver, the sound signal must be attenuated by about 20 dB in the filter with respect to the peak of the filter characteristic, and very accurately located minima ('traps') must be included in the filter response to coincide with the carrier frequencies of adjacent channels. The specification for the amplitude response for a bandpass filter for the U.K. television system is shown in *fig. 13*.

The requirements for the centre frequency and bandwidth of this filter are such that the component is ideally suited to realization by acoustic surface-wave techniques. The substrate size is reasonably small, yet the photolithographic definition required is within the capability of current IC technology. A conventional television i.f. filter incorporates a number of inductors that have to be individually aligned in the factory. The potential advantages of surface-wave filters — no adjustment, good reproducibility and good reliability — have obvious attractions.

The cost of a device intended for a very large consumer market is as important as its technical performance, and this fact has significantly influenced the direction of our research. The first, and perhaps the most important, decision to be made before making a filter is what to use for the substrate material. As explained earlier, the frequency response of the fllter depends directly on the surface-wave velocity v, and the material must therefore have a highly reproducible value of v and one that remains sufficiently constant with temperature and time. This is particularly important for the trap frequencies, which must coincide accurately with the carrier frequencies of adjacent channels. The permissible difference in v from filter to filter is 0.25 % at most, the temperature coefficient must

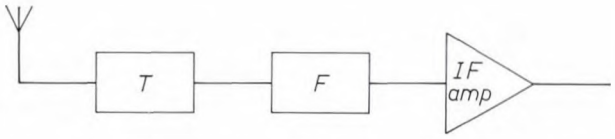


Fig. 12. In a television receiver the intermediate-frequency filter (F) is connected between the tuner (T) and the i.f. amplifier (*IF amp*).

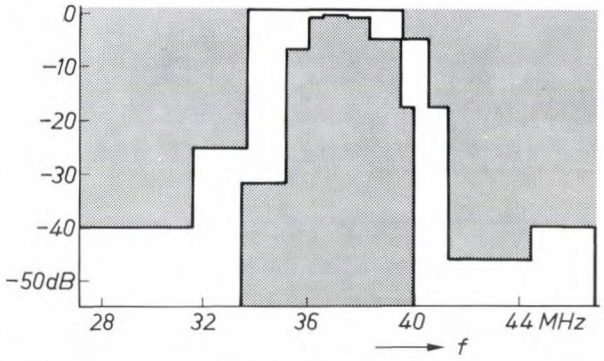


Fig. 13. Tolerance mask for the amplitude response of a television filter for the system in use in Britain (System I). The amplitude response must fit between the hatched regions.

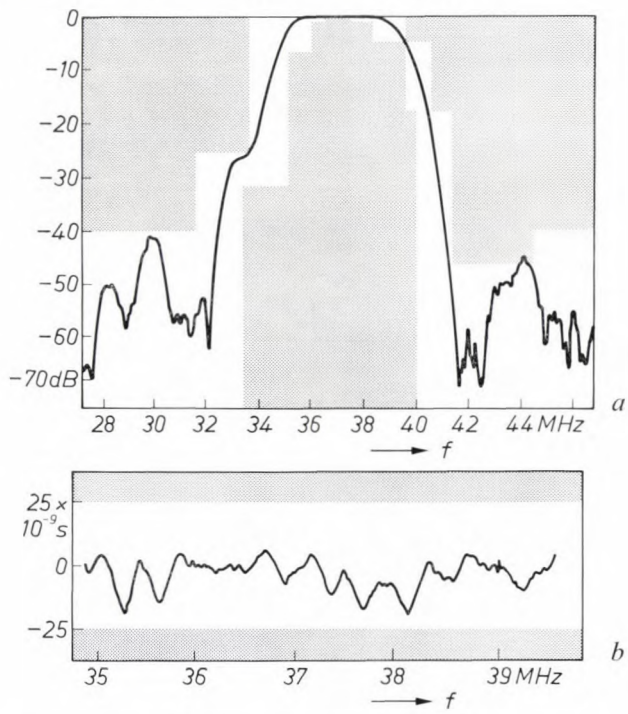


Fig. 14. *a*) As fig. 13, but now including the amplitude response of one of our filters. This completely satisfies the specification. *b*) The group-delay response of this filter, again with the appropriate tolerance mask.

be smaller than $120 \times 10^{-6}/°C$ and ageing drift during the life of the filter (about 10 years) must be less than 0.5 %.

The most commonly used material for surface-wave filters is lithium niobate, $LiNbO_3$, usually with propagation in the z-direction in a y-plane. The value of k^2 is 0.048, which is very high. However, we considered that this material was too expensive for use in a tele-

vision filter. In the initial phases of our work we considered the use of piezoelectric ceramic materials, because of their low cost, but we found that the reproducibility of the surface-wave velocity v was only about 1%, which was far too large to be acceptable. We therefore returned to the single-crystal materials, which are more than adequate for the reproducibility requirements.

One material that we have investigated is bismuth silicon oxide, $Bi_{12}SiO_{20}$, and most of our work has been concentrated on filters with this material as the substrate [14]. It should be a cheaper material to grow than $LiNbO_3$, since it has a lower melting point and is made from cheaper raw materials. It is not ferroelectric and so does not require polarization. The surface-wave velocity is about half that of lithium niobate, which means that the filters are much smaller, so that more can be made from a slice of given size, thus reducing manufacturing costs. In the direction that we use (a $\langle 110 \rangle$ direction in a (100) plane), it has an acceptable value of k^2 (0.018) and a temperature coefficient of delay of $118 \times 10^{-6}/°C$.

The insertion loss of the i.f. television filter must not be too high, or the signal-to-noise ratio of the receiver will be too small. It is difficult to compensate for this by extra gain in the tuner, because of signal-handling problems. On the other hand, the insertion loss of a surface-wave filter should not be too low either, because of the triple-transit signals. If these reflected signals are not at least 40 dB below the main signal, they will produce 'ghost' images in the television picture. As mentioned earlier, the triple-transit signal increases in strength as the insertion loss decreases, and its level is about twice the insertion loss plus an extra 6 dB. If the triple-transit signal is at the stated level of —40 dB, the insertion loss of the surface-wave filter is about 17 dB. This situation would be altered if any of the schemes referred to earlier for reducing the triple-transit signal were to be used, but in the case of the television filter the associated costs would be too great. An insertion loss of 17 dB is obtained from a filter on a substrate of $Bi_{12}SiO_{20}$ with a low-Q tuning inductor at one end; television receivers using such a system have been found to give a very satisfactory performance.

The details of the frequency-response requirements are not the same in the various European countries or in other parts of the world. So far we have concentrated on producing designs for System I (U.K.) and System G (the Netherlands, West Germany, etc.). The amplitude response of a filter for System I is shown in *fig. 14a*. This filter is a single-finger design; the uniform output transducer has 16 fingers and the input transducer 74. The substrate material is $Bi_{12}SiO_{20}$. The 'low-interaction' layout mentioned earlier is used in the input transducer, and this helps to keep the group-delay response flat to within 25 ns across the passband, as shown in fig. 14*b*.

The other system (G) is significantly different from the British one. Here the signal is transmitted with a predistorted group-delay characteristic. This is done to compensate for the group-delay characteristics of conventional inductance-capacitance filters. This means however that the surface-wave filter must accurately reproduce this group-delay distortion. The amplitude and group-delay requirements are shown in *fig. 15*, together with the measured response of one of our surface-wave filters. This filter was also made on $Bi_{12}SiO_{20}$ and has a uniform output transducer with

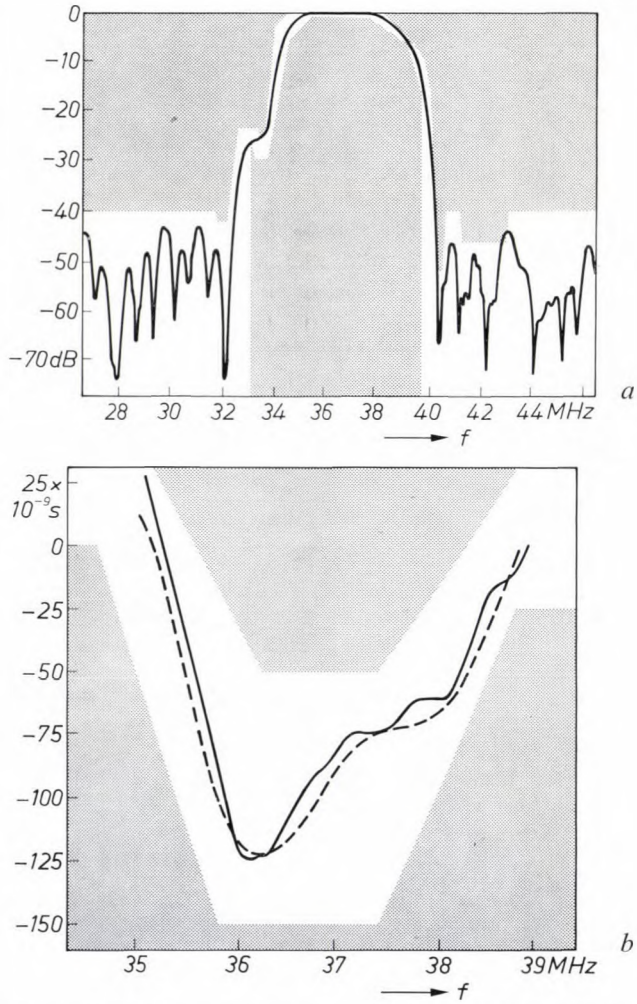


Fig. 15. *a*) Tolerance mask and measured amplitude response of a television filter for system G (the Netherlands, West Germany, etc.). *b*) Tolerance mask and measured group-delay response for a filter for system G. This response is designed in such a way that group-delay distortion occurring elsewhere in the system is compensated. The dashed curve is the calculated response.

[14] Another team at our laboratories have developed methods for making large single crystals of this material; see J. C. Brice, T. M. Bruton, O. F. Hill and P. A. C. Whiffin, The Czochralski growth of $Bi_{12}SiO_{20}$ crystals, J. Crystal Growth **24/25**, 429-431, 1974.

18 fingers. In this case a 200-finger double-finger structure was used for the input transducer. As can be seen in fig. 15, both the amplitude and the group-delay characteristics lie within the tolerances. There is also good agreement between the predicted group-delay characteristic and the measured one as shown in fig. 15*b*. The mask used to make this filter is shown in *fig. 16*.

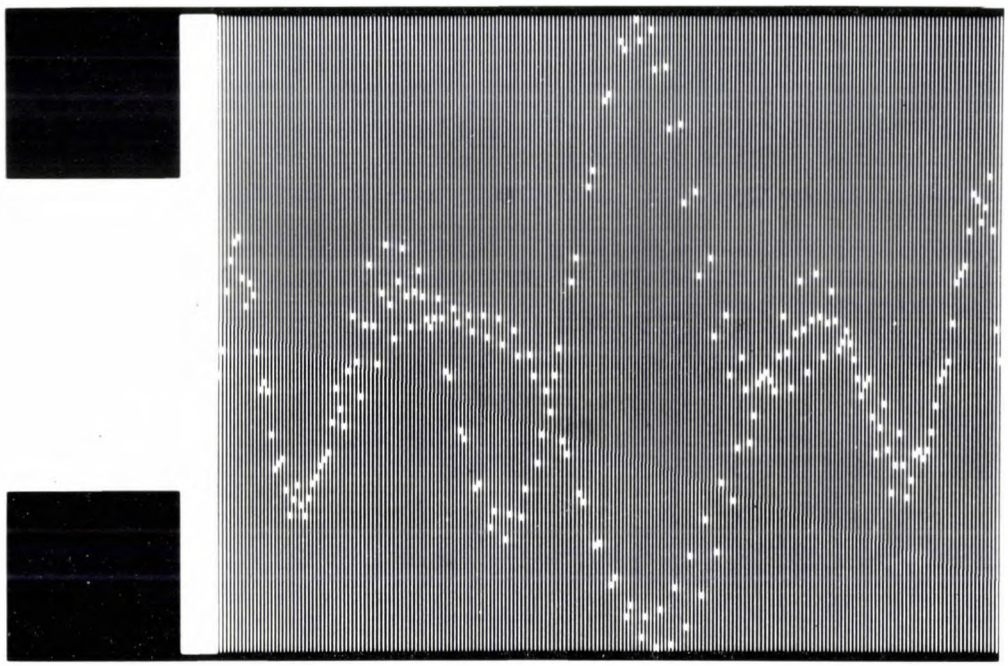


Fig. 16. A part of the mask used for making the transducers in the filter of fig. 15. This part relates to the input transducer, which has 200 fingers.

Other filters

Surface acoustic-wave filters have been designed for other applications, such as i.f. filters for radar and telecommunication. Here the cost of the filter is not such an important factor as with the television filter.

In radar, where the systems deal with pulses, the shape of the signals produced by the filter is no less important than the frequency response. This means in particular that the triple-transit echo of the surface-wave filter must be reduced to a level below —60 dB. The insertion loss of a radar filter is usually designed to be between 25 and 30 dB so that the triple-transit echo level is suitably low; occasionally special transducer arrangements are used that give cancellation or partial cancellation of this echo. We shall now discuss three applications of the surface-wave filter in radar and telecommunication.

Gaussian filters

One type of response that can easily be produced with a surface-wave filter is the Gaussian function. This response is of particular interest to radar engineers, as it is a shape that gives no ringing of the input pulse. Since in this application the acceptable insertion loss is fairly high, we have constructed such a filter on a $Bi_{12}SiO_{20}$ substrate with propagation of the waves in a $\langle 100 \rangle$ direction on a (110) plane. This orientation is particularly favourable for the production of surface waves without the simultaneous production of bulk waves [10]. The response of such a filter with a bandwidth of 4 MHz (measured between the 3-dB points) is shown in *fig. 17*; the central part of the filter characteristic is compared with the theoretical curve in *fig. 18*. The excellent stop-band responses are primarily due to the almost complete elimination of bulk-wave responses.

Matched filters

A class of filters that are eminently suitable for realization as surface-wave filters is that of the matched filters. These filters are used for the optimum detection of pulses in the presence of noise. This requires that the frequency response of the filter is matched to the fre-

quency spectrum of the received pulse. For example, the frequency spectrum of a rectangular pulse is of the form $(\sin x)/x$; for the optimum detection of this pulse in the presence of Gaussian noise, we require a filter with a triangular function at the output as shown in *fig. 20*.

The improvement in signal-to-noise ratio over the corresponding filter using only the central lobe of the

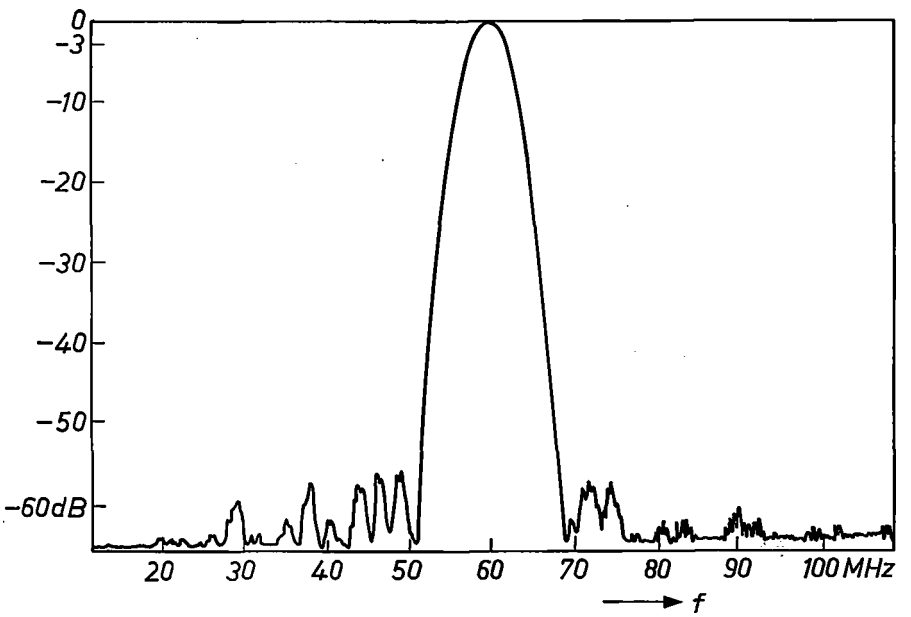


Fig. 17. Gaussian response obtained with a surface-wave filter. Filters with a Gaussian response are applied in radar systems, since they do not introduce 'ringing'.

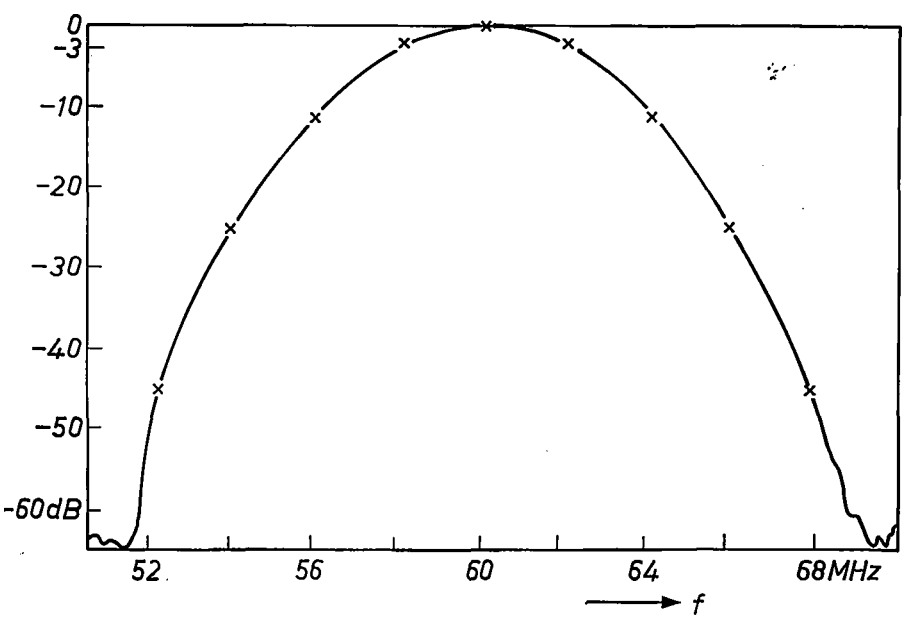


Fig. 18. The central part of the response of fig. 17. The crosses are points from the theoretical Gaussian response.

with an identical frequency response. For a pulse with a duration of 10 μs, modulating a carrier of 13.5 MHz, the $(\sin x)/x$-like response is shown in *fig. 19*. The matched filter responds to the rectangular pulse response — which approximates to the usual i.f. response — is 2.5 dB. Matched filters can be applied with effect only if the signal to be detected has a well defined frequency spectrum. Unfortunately, in radars

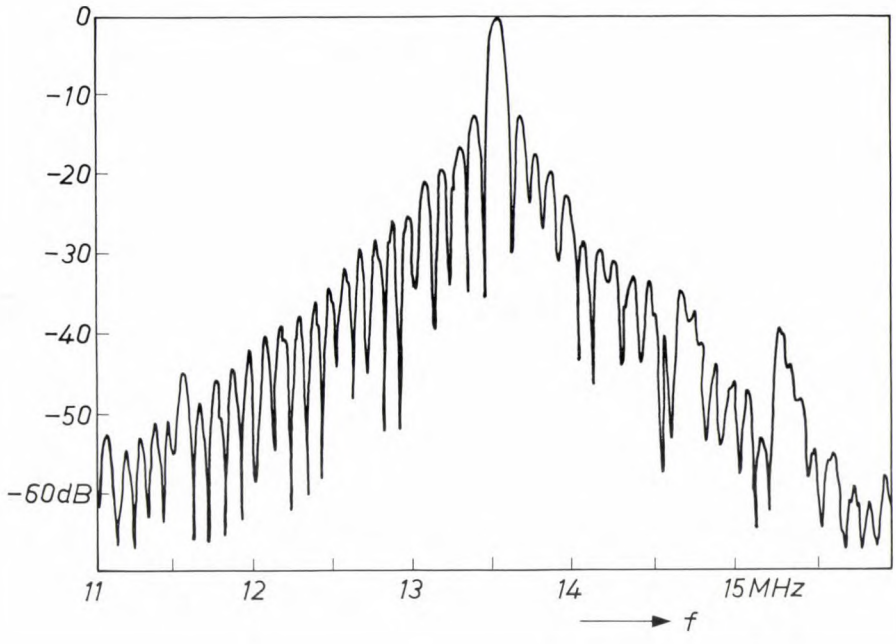


Fig. 19. Frequency characteristic of a matched filter. The filter is matched to the spectrum of 10-μs rectangular pulses used to modulate a 13.5-MHz carrier. Such a filter gives better separation of signal from noise. (The improvement with respect to a conventional filter is 2.5 dB.) For rectangular pulses the frequency response should have the form $(\sin x)/x$.

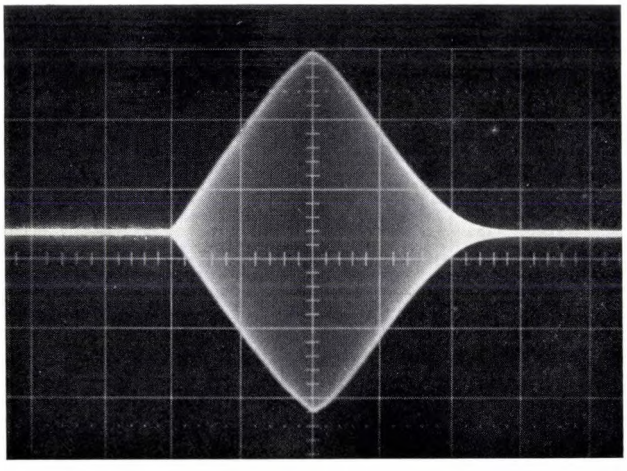


Fig. 20. Output signal from the filter of fig. 19 when a pulse of carrier frequency 13.5 MHz is applied to the input to the filter.

that use magnetrons as the power source the spectrum is too ill-defined for this method to be used, and it is therefore limited to the more advanced types of radar systems.

In some situations a filter whose response is an approximation to that of a matched filter will give adequate results. As an example, *fig. 21* shows a $(\sin x)/x$ response that has been limited to the central lobe and the first two side lobes. Such a response would

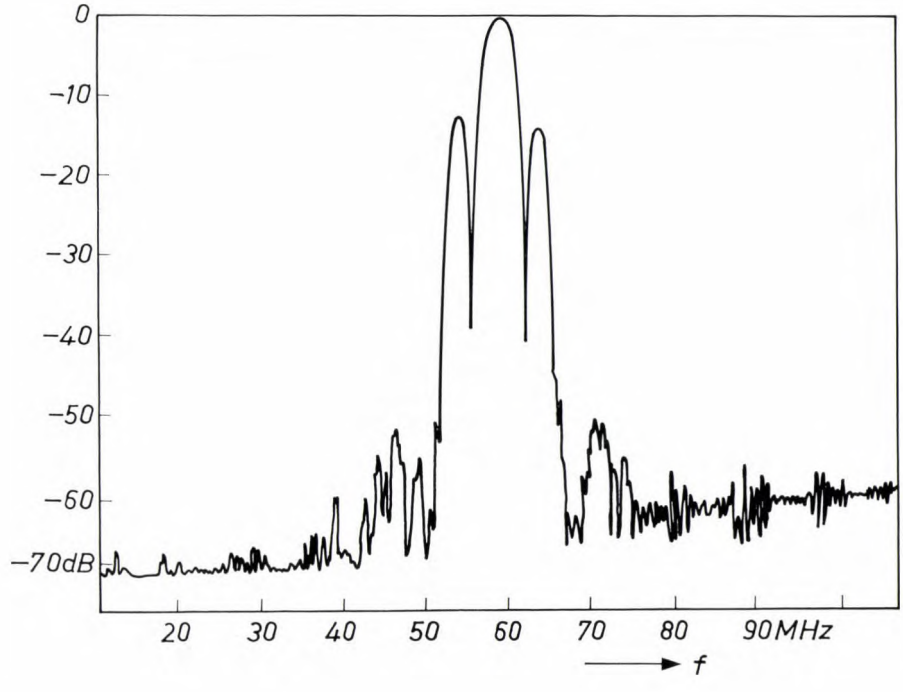


Fig. 21. Approximation to the matched filter. The frequency response consists of the central part of the function $(\sin x)/x$ with the first two side lobes. This response was also obtained with a surface-wave filter. The improvement with respect to a conventional filter is about 1 dB.

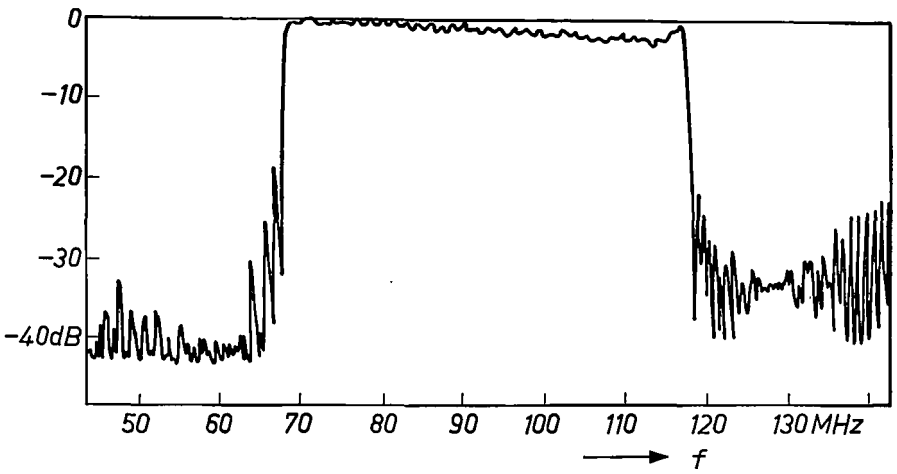


Fig. 22. Amplitude response of a wideband filter that is virtually flat from 65 to 115 MHz.

give an improvement of about 1 dB in signal-to-noise ratio, compared with that of a conventional filter. It is difficult to think of any other way of producing such a response without the use of surface waves.

Wideband filters

In telecommunication systems filters are often necessary that will have a flat frequency response over a wide frequency band. Such filters, with relative bandwidths of up to 50%, have been realized with surface-wave techniques. To keep the insertion loss to reasonable levels, they were constructed on $LiNbO_3$, so as to take advantage of the high value of k^2. These filters are usually tuned at both ends with a series inductor; the effect of this tuning on the shape of the response must be taken into account in the design procedure. An example of such a filter is shown in *fig. 22*.

The work on filters for radar and telecommunication was carried out with the support of the Procurement Executive, Ministry of Defence U.K., sponsored by CVD.

Summary. Acoustic surface waves on a substrate of piezoelectric material can be employed in an analogue bandpass filter. The frequency response of the filter is determined by the form and dimensions of the interdigital electrodes in the transmitting and receiving transducers. A relatively simple model provides a means of designing these filters in a systematic way. However, this model requires correction for factors such as diffraction, reflection, and interactions between the acoustic sources. Computer programs have been developed that design a layout for a mask-making system, with the specifications for amplitude and group-delay characteristics as the input data. The article considers, as an application, an analogue bandpass filter for use as an i.f. filter in television receivers. The devices have been developed to a point at which they can be considered for commercial production in quantity. Good results have also been obtained with the application of surface-wave techniques to Gaussian filters, matched filters and wideband filters.

An automated Langmuir trough for building monomolecular layers

The number of methods available for arranging molecules in a particular order, other than by crystallization, is very limited. One of them is the Langmuir and Blodgett method [1], in which monomolecular layers of a surface-active substance (or 'surfactant'; for example a fatty acid or an ester) on a liquid surface are transferred to a substrate (e.g. of glass) by drawing a strip of substrate upright through the surface layer (*fig. 1*). This method makes it a relatively simple matter to build many tens of monomolecular layers of different chemical composition in any desired sequence [2].

We have built an instrument that works on this principle and is free of the difficulties encountered with other equipment. The design was aimed at functional mechanization and automatic control. The instrument (*fig. 2*) consists of a shallow trough made of a chemically inert synthetic material, which is filled with water or an aqueous solution carefully freed from surface contamination. The trough is filled 'brimfull', i.e. until the liquid level is just *above* the edge of the trough. The desired monomolecular layer is then applied to the water surface. When the apparatus is in operation, a moving barrier, like a kind of piston, gradually pushes the molecules of the surfactant along the surface. Appropriate geometry and measures such as the use of a material not readily wetted by water ensure that the surfactant does not leak away along the barrier. The barrier maintains a constant surface pressure in the layer, so that monomolecular layer to be transferred remains homogeneous. Transfer is effected by means of an immersion mechanism that moves one or two strips of the chosen substrate up and down through the surface layer. The barrier is set in motion by a toothed belt driven by a stepping motor, which is controlled by the surface pressure in the monomolecular layer. This pressure is measured by the Wilhelmy method (1863) [3] with a balance that delivers the electrical control signal.

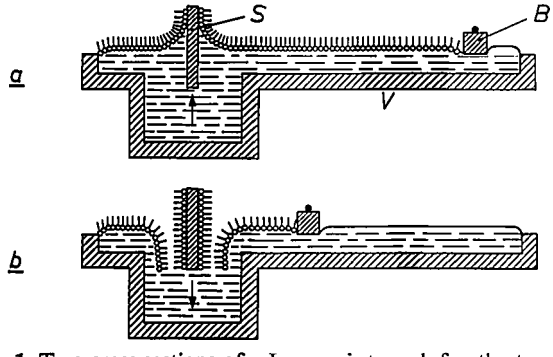


Fig. 1. Two cross-sections of a Langmuir trough for the transfer of monomolecular layers. The trough V contains a salt-water solution on top of which there is a monomolecular layer of a substance that lowers the surface tension, e.g. a fatty acid. The substance is most easily 'spread out' by applying it to the water as a solution in benzene or chloroform; the solvent then evaporates. The polar 'heads' (the COOH groups) of the fatty-acid molecules are located in the water. The aliphatic hydrocarbon chains of the molecules are hydrophobic and stick into the air. The thickness of the layer is equal to the length of the molecules, which proves that the tails stand upright. By slowly pulling up a substrate (S) through the surface layer it is possible to transfer to it a monolayer (a 'Langmuir-Blodgett' layer). A second layer can be placed on the first ('tail to tail') by moving the strip down and then up again, and so on. In the trough described here a barrier B keeps the surface pressure in the layer constant, ensuring a 'neat' arrangement of the molecules. The surface pressure is kept constant by a control mechanism that ensures that the barrier moves along at a controlled rate during the transfer to S.

In the Wilhelmy method a strip of filter paper is suspended vertically from the arm of a balance and is partly immersed in the liquid. A change $\Delta\pi$ in the surface pressure gives rise to a change ΔF in the force acting on the beam, given by:

$$\Delta F = -s\,\Delta\pi,$$

where s is the total circumference (in the liquid surface) of the filter paper. If the material of the strip does not permit complete wetting, a factor $\cos\phi$ must be added to the right-hand side of the equation. The angle ϕ is the contact angle between the material and the liquid (*fig. 3*). In the case of filter paper, which is completely wetted, ϕ is equal to zero.

The control system keeps the surface pressure at any desired value, with a relative deviation of no more than 1%. In most cases the speed of response to pressure changes is determined by the viscosity of the monomolecular layer, and not by the electrical and electromechanical parts of the system.

Finally, there is also a control mechanism that keeps the liquid level constant (to within 30 μm). This control is needed first and foremost to prevent leakage along the barrier, and it also improves the accuracy with which the surface pressure is measured. The control mechanism consists of a float that shifts the ferrite core of a differential transformer when it is displaced, thereby opening or closing a magnetic valve in the tube connecting the trough and a liquid reservoir.

The trough is mounted in a glove box in which temperature and relative humidity are kept constant. To minimize vibration the complete system stands on rubber strips on a heavy table.

The main difference between the new instrument and the classical version is the presence of the leakproof

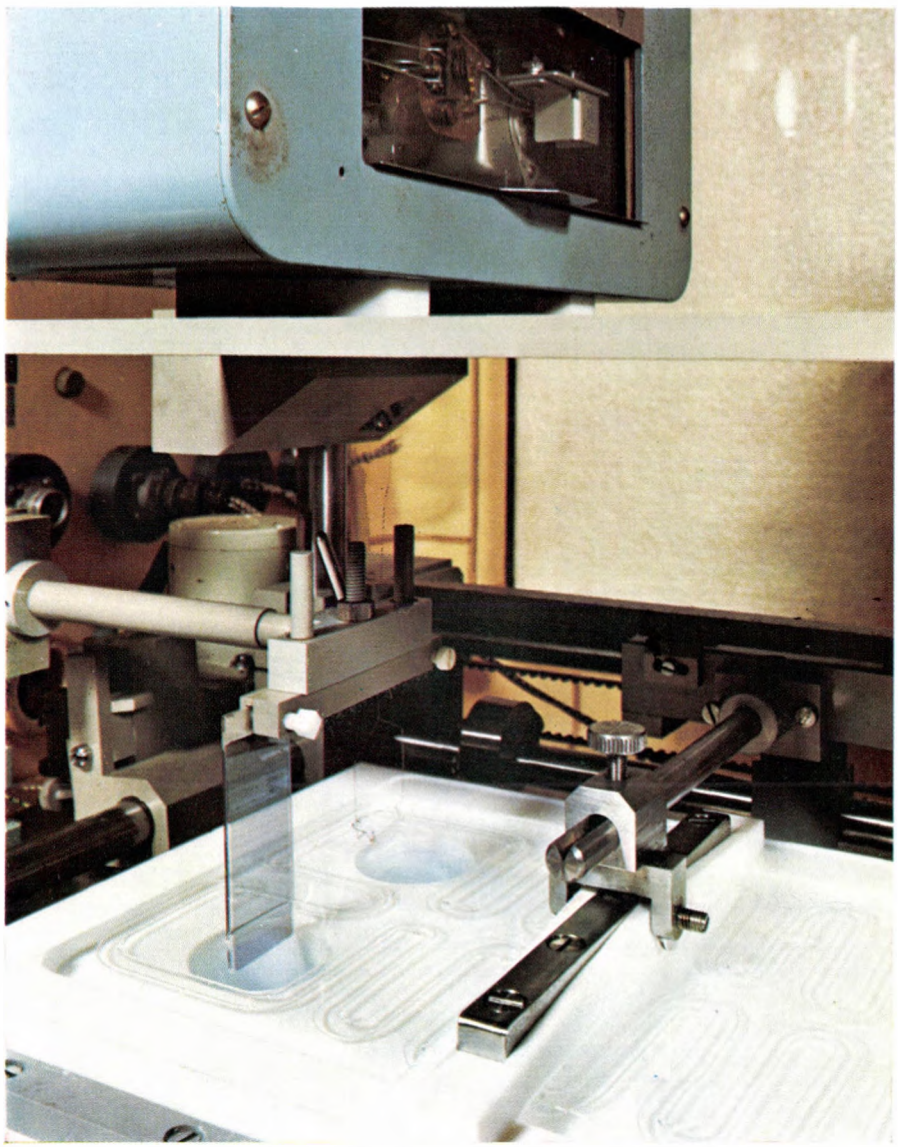


Fig. 2. The Langmuir trough with automatic pressure control. The trough is made of white synthetic material with inside dimensions of $50 \times 15 \times 1.5$ cm, and has two sunken parts to allow sufficient room for the immersion movement. Part of the immersion mechanism, which can move one or two substrate strips up and down, can be seen on the left. The movable barrier is at the centre. In the left background there is a strip of filter paper, suspended from the arm of a balance, which is partly immersed in the liquid and is used for measuring the surface pressure. The right-hand part of the trough, which is not visible, contains a level controller. The tubes near the bottom of the trough belong to the thermostat that holds the water temperature constant.

barrier. The pressure in the layer to be transferred was formerly produced and kept constant by a film of oil ('piston oil') that was also applied to the surface and separated from the surfactant by a waxed thread. In this method the surfactant could easily become contaminated by the oil. With such an oil, of course, the surface pressure could only be fixed at one value; with the new instrument there is a free choice. A *floating* barrier has also been used in earlier methods [4].

As an example of work done with the Langmuir trough discussed here, *figs. 4*, *5* and *6* present some results of an investigation into the influence of the structure of the dispersed layer on the deposition ratio [5], i.e. the ratio between the quantities of transferred material for upward and downward movement

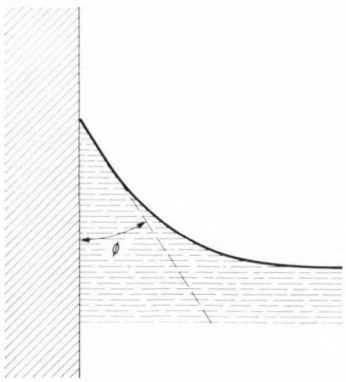


Fig. 3. The contact angle ϕ between a strip of solid material (left) and a liquid. For complete wetting the contact angle is zero.

[1] Katharine B. Blodgett, J. Amer. Chem. Soc. **57**, 1007, 1935.
[2] H. Kuhn, D. Möbius and H. Bücher, in: A. Weissberger and B. W. Rossiter (eds), Physical methods of chemistry, Wiley, New York 1972, Part I, p. 577.
[3] See for example G. L. Gaines, Jr., Insoluble monolayers at liquid-gas interfaces, Interscience, New York 1966, p. 45 ff.
[4] I. H. Sher and J. D. Chanley, Rev. sci. Instr. **26**, 266, 1955.
[5] E. P. Honig, J. H. Th. Hengst and D. den Engelsen, J. Colloid Interf. Sci. **45**, 92, 1973.

of the strip of substrate material. The surfactant used was a mixture of arachidic acid ($C_{19}H_{39}COOH$) and its methyl ester. The deposition ratio of this mixture depends on its composition (fig. 4). For the same mixtures the surface pressure in the monomolecular layer is measured as a function of the average surface area per molecule (fig. 5). The average surface area can be calculated from the position of the barrier and the quantity of surfactant on the water. The extremely steep part of the curves is called the 'solid phase' of the surface layer (very small changes in area are associated with very marked changes of pressure). The transition region around the knee in the curves resembles the situation in a liquid. For higher values of the average surface area per molecule the surface pressure is almost zero: the surface-active effect of the mixture is then negligible. If the surface pressure at the transition point between 'solid phase' and 'liquid phase' is plotted as a function of the composition of the mixture (fig. 6), we see that there is a striking correspondence with fig. 4. This indicates that the deposition ratio is determined by molecular interactions in the monomolecular layer.

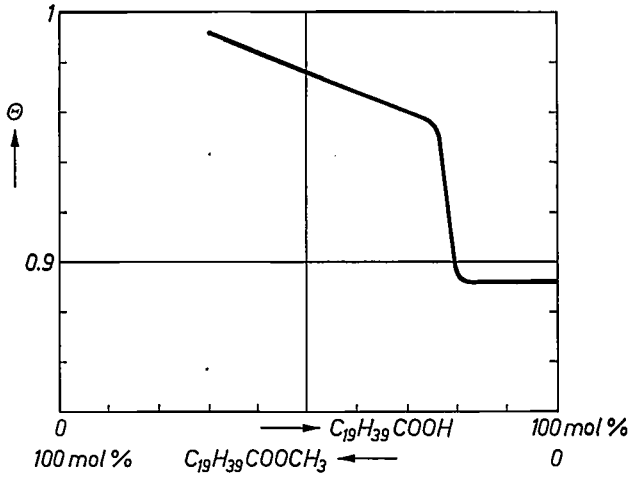


Fig. 4. The measured deposition ratio Θ — the ratio between the quantities of material deposited during upward and downward movements — during the building of monomolecular layers consisting of a mixture of arachidic acid and its methyl ester, as a function of the composition of the mixture. The trough was filled with an aqueous solution of cadmium chloride.

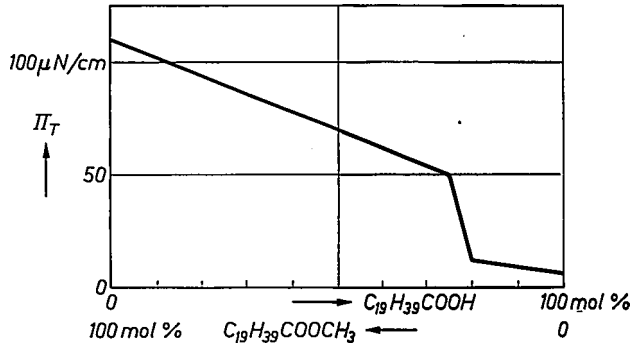


Fig. 6. The surface pressure Π_T at the knee points in the curves of fig. 5, plotted as a function of the composition of the mixture. There is a striking correlation between this curve and that of fig. 4, indicating a correlation between the surface pressure at the knee point (a structure parameter of the layer) and the deposition ratio during transfer.

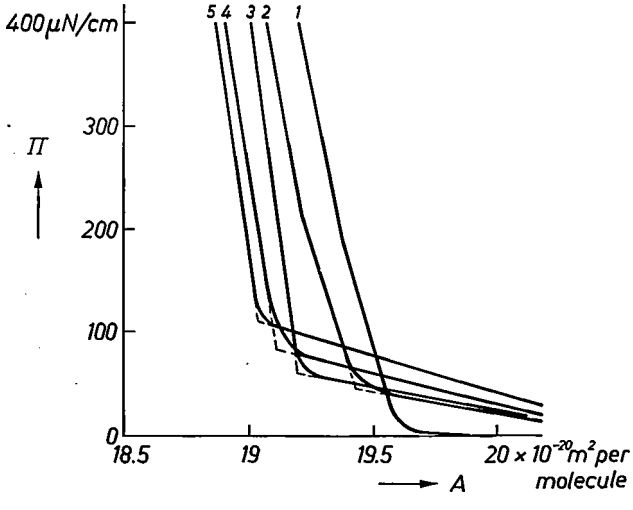


Fig. 5. The measured surface pressure Π of a monomolecular layer consisting of a mixture of arachidic acid and its methyl ester, as a function of the average surface area A per molecule. *1.* pure acid. *2.* acid 75% (mol.), ester 25%. *3.* acid 50%, ester 50%. *4.* acid 25%, ester 75%. *5.* pure ester. To the right of a knee point the surface-active effect of the mixture is weak. To the left of that knee point the monolayer is said to be in a 'solid phase'.

D. den Engelsen
J. H. Th. Hengst
E. P. Honig

Dr D. den Engelsen, J. H. Th. Hengst and Dr E. P. Honig are with Philips Research Laboratories, Eindhoven.

Light transmission of sintered alumina

J. G. J. Peelen

Sintered alumina is the material most widely used for making the discharge tubes of high-pressure sodium lamps, because it combines good light transmission with the required chemical resistance to sodium vapour at 1500 K. Depending on the nature of the sintering process, the material either scatters light very little — it is 'transparent' — or it scatters light very strongly — it is only 'translucent'. Partly because it seemed to be possible to improve the light transmission — in which respect the transparent material and the translucent one are almost on a par — a closer investigation has been made of the correlation between these properties.

Sintered alumina for sodium lamps

The discharge tube of the high-pressure sodium lamp, familiar for its golden-coloured light, is almost invariably made of sintered aluminium oxide (alumina). In addition to giving reasonable light transmission, this polycrystalline material is also exceptionally resistant to the hot sodium vapour present in lamps of this type. The temperature of the coldest spot is no less than about 1000 K, and between the electrodes the temperature can be another five hundred degrees higher. Ordinary types of glass and even fused silica cannot be used under such conditions. As well as good chemical resistance to sodium vapour at 1500 K, the discharge-tube material is required to have a low vapour pressure, high mechanical strength and electrical resistance, and of course the light loss in the visible part of the spectrum must be as low as possible.

The use of ceramics was envisaged right from the beginning, because of their chemical resistance and exceptionally good refractory properties. The light loss caused by such materials when used in lamp envelopes, however, is a troublesome complication. To overcome this difficulty it is necessary to know the relationship between the microstructure of the material and its optical properties.

As long ago as the fifties it was considered that polycrystalline alumina (Al_2O_3) could satisfy the requirements of an envelope material if the density of a sintered-aluminina tube could be increased to a value close to 100%. This view was based on the knowledge that absorption and reflection are of little significance in this material, and it seemed that light scattering by pores left behind after sintering was the main reason for the light loss. Some of the incident light would then be back-scattered from the material, and some of the rest would travel such a long path owing to repeated scattering inside the material that absorption would become significant. The light would therefore emerge from the material considerably attenuated in the forward direction. To avoid this light loss a material of very low porosity must therefore be used.

At the time, the problem in sintering alumina was the effect of discontinuous grain growth [1]. The somewhat larger grains, with more strongly curved interfaces, tended to grow very fast, so that pores became enclosed in the grains and were unable to diffuse out. R. L. Coble [2] succeeded in suppressing discontinuous grain growth by adding 0.2 per cent by weight of MgO to the alumina. This reduces the number of pores that are trapped inside the grains, and prolonged sintering then produces a material that transmits light reasonably well, though very diffusely. This material has found wide application; the lamp envelopes made with it proved to be capable of withstanding attack from the hot sodium vapour for many thousands of hours.

This was of course an important practical result, but it did not answer the questions of how the light transmission is related to the microstructure and of how this transmission might be improved by other changes in the sintering process. This relationship has been the subject of extensive investigations at Philips Research Laboratories. They showed, for instance [3], that 'continuous hot pressing' [4] at a relatively low tem-

Drs J. G. J. Peelen is with Philips Research Laboratories, Eindhoven.

[1] See A. L. Stuijts, Philips tech. Rev. **31**, 44, 1970.
[2] R. L. Coble, J. appl. Phys. **32**, 787 and 793, 1961.
[3] J. G. J. Peelen, Science of Ceramics **6**, XVII, 1973.
J. G. J. Peelen and R. Metselaar, J. appl. Phys. **45**, 216, 1974.
[4] G. J. Oudemans, Proc. Brit. Ceram. Soc. **12**, 83, 1969.

perature resulted in a product that scattered light much less than the one mentioned above, a result for which there was no immediate physical explanation.

In common parlance the hot-pressed material would be described as 'transparent', as opposed to the other type that is better described as 'translucent' (*fig. 1*).

This article will mainly be concerned with the optical aspects of the investigation. Among the questions studied was the extent to which the transmitted light, apart from being scattered by pores, is also affected by scattering and birefringence at grain boundaries. This effect has been found to be virtually negligible. The calculations of light scattering by pores [*] were based on Mie's equations for light scattering by spherical particles [5]. The parameters in these calculations are the porosity of the material and the characteristic data for the distribution of the pore radii. Comparison of the measured transmission spectra from samples of both 'transparent' and 'translucent' material with calculated spectra allows these quantities to be determined; this rather tends to confirm our views concerning the light transmission in polycrystalline alumina.

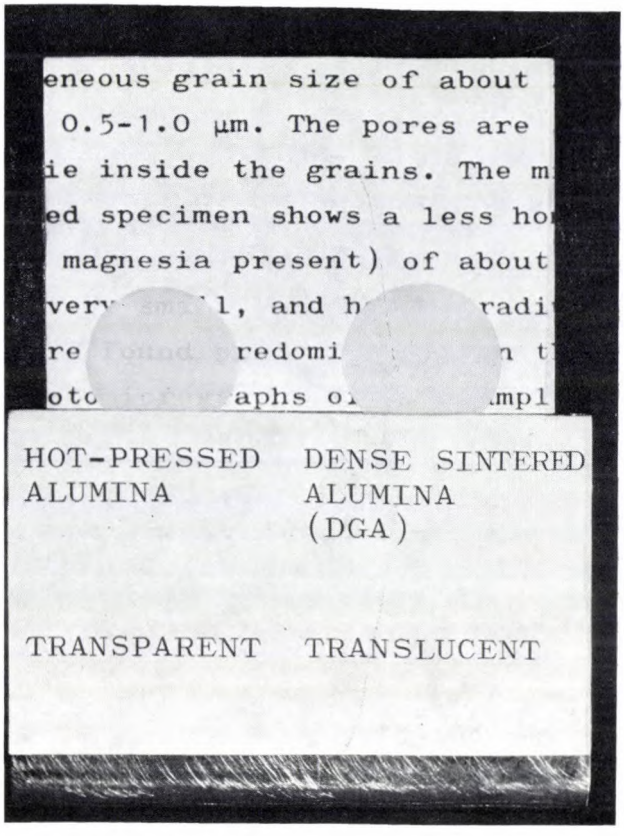


Fig. 1. The photograph shows two discs of polycrystalline Al_2O_3, about 10 cm above a printed page. One was made by continuous hot pressing, the other was sintered with the addition of a small quantity of MgO. The text under the hot-pressed disc is legible since this material causes very little light scattering (see fig. 4).

G. Mie's treatment of light scattering considers a plane electromagnetic wave incident on a spherical body. The scattered waves are calculated from Maxwell's equations by requiring that the waves at the surface of the sphere satisfy certain boundary conditions. The solution contains the wavelength λ in the medium surrounding the sphere, the normalized radius a/λ of the sphere and the relative refractive index $n_{\text{sphere}}/n_{\text{medium}}$. The amplitude of the scattered waves — which depends on the direction — is found to be the sum of a series of Bessel functions with the normalized radius in the argument, each multiplied by a Legendre polynomial. The direction-dependence is expressed in these polynomials. Three situations may be distinguished: $a < \lambda$, $a \approx \lambda$ and $a > \lambda$. In the first case the first term of the series is sufficient; this term represents the Rayleigh scattering. For the case $a \approx \lambda$ no simple approximations are known. In the case $a > \lambda$ there tends to be diffraction rather than scattering, and approximate formulae for this do exist. In our case the most common situation is $a \approx \lambda$, which means that the scattered light can only be determined by computer calculations.

Path of rays and transmission

Fig. 2 shows the situation that arises when a narrow beam of light passes through a layer of polycrystalline alumina. We shall take the ratio I/I_0 of the light intensities in the direction of incidence as a measure of the transmission. In view of the scattering this in-line transmission is a minimum measure of the light transmission of the layer.

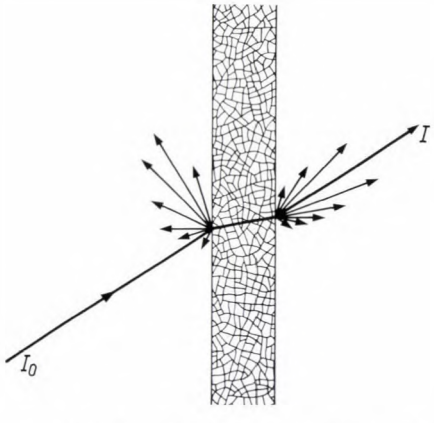


Fig. 2. Paths of rays found when a narrow beam of light passes through a layer of porous polycrystalline alumina. Scattering in the material reduces the transmission in the direction of the incident beam, resulting in diffuse reflection and diffuse transmission. The normal reflection from the front and back of the layer is not shown; the associated transmission loss is about 15%.

The light intensity — the luminous flux per solid angle — was measured with a spectrophotometer which records the luminous flux in a cone of very small apex angle (about 0.02 rad). Since this apex angle is not equal to zero, scattered waves are measured in addition to the unaffected waves in the direction of incidence. Some transmission curves determined in this way are shown in *fig. 3*. The exceptional steepness of the curve for hot-pressed alumina indicates a Rayleigh-type scattering, whose effect is inversely proportional to the

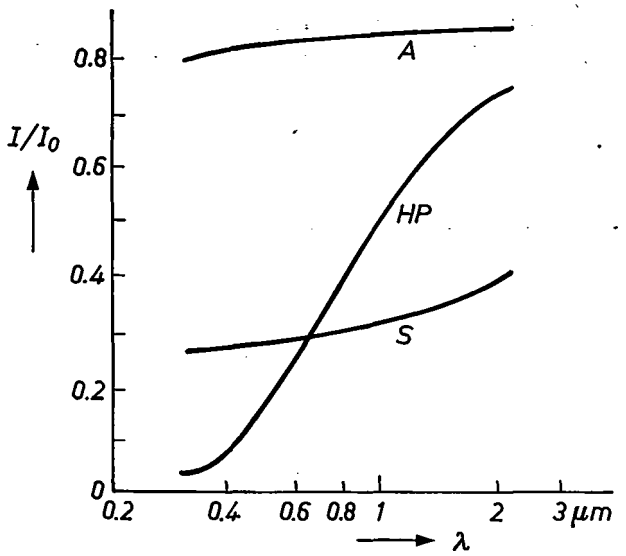


Fig. 3. Measured transmission I/I_0 of sintered (*S*) and hot-pressed (*HP*) alumina as a function of the wavelength of the light. The term transmission refers to the ratio of the light intensities taken in the direction of incidence; diffuse light is therefore not included in the measurements (fig. 2). The thickness d of the layer is 0.5 mm. The porosity V_p is 1×10^{-3} (volume fraction). For comparison the curve for sapphire is shown (*A*), the single-crystal form of Al_2O_3.

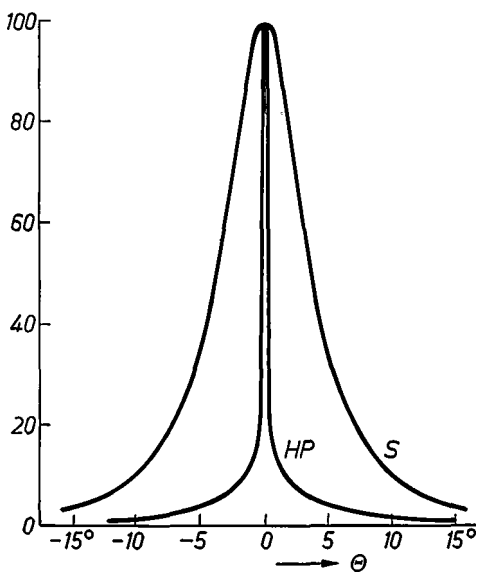


Fig. 4. Scattering profile of transmitted light ($\lambda = 0.546$ μm), measured for sintered (*S*) and for hot-pressed (*HP*) alumina. The relative light intensity is plotted vertically as a function of the angle Θ between the ray considered and the direction of incidence. The light intensity in the direction of incidence is arbitrarily assumed to be 100 in both cases. The narrowest profile belongs to the material that did not affect the legibility of the text in fig. 1.

fourth power of the wavelength. A second example of the experimental side of our investigation is the curve shown in *fig. 4*, representing the scattering profile of the transmitted light for both conventionally sintered and hot-pressed alumina [6]. It can be seen that the light propagates far less diffusely in the hot-pressed material, while at the same time the transmission is less — at the wavelength used. The last, perhaps unexpected, result is equally understandable when Rayleigh scattering dominates, since this is also strong in the backward direction.

As noted above, the transmission loss is mainly attributable to scattering in the material. If the light travels distances that are no greater than 0.5 to 1 mm, absorption is ruled out because the absorption coefficient of the material is extremely small. Reflection from the front and back of the layer (fig. 2) does, however, make some contribution to the transmission loss. When the light is normally incident on a completely smooth surface, the fraction that is not transmitted because of reflection is about 15%. The extent to which the transmission is affected by the birefringence at the grain interfaces of the material depends on the number of interfaces and thus on the size of the grains. An exact calculation of this is difficult, but a kind of overestimate can be made, by assuming the maximum effect on the transmission at each new refraction. For a material thickness corresponding to 50 grains, the transmission loss estimated in this way is completely negligible [7].

The scattering of the light in the material is fundamentally due to regions of differing refractive index, such as pores or particles of a second phase — whose presence is due to dopants or impurities — and also, apparently, to grain boundaries. The grain boundaries come into the picture because particles of a second phase or certain segregations prefer to concentrate there; in themselves the grain boundaries cause no light loss of any significance. This statement is based on the observation that certain single-crystal materials — which contain no grain boundaries of course — show no change whatever in their optical properties in the visible part of the spectrum on a transition to a polycrystalline form. The region of atomic disorder at the grain boundary is extremely small, perhaps no more than a few thousandths of the average wavelength in the visible region. Rayleigh scattering is certainly possible here, but the number of such scattering centres in the light path is so small that no observable effect is to be expected.

This leaves us with pores and possible particles of a second phase as the centres that scatter the light. In

[*] These calculations were performed in cooperation with Dr R. Metselaar of these laboratories.

[5] H. C. van de Hulst, Light scattering by small particles, Wiley, New York 1957, chapter 9.
[6] J. G. J. Peelen, Conf. on Microstructure of Ceramics, Oxford, April 1975.
[7] N. Grimm, G. E. Scott and J. D. Sibold, Amer. Ceram. Soc. Bull. 50, 962, 1971.

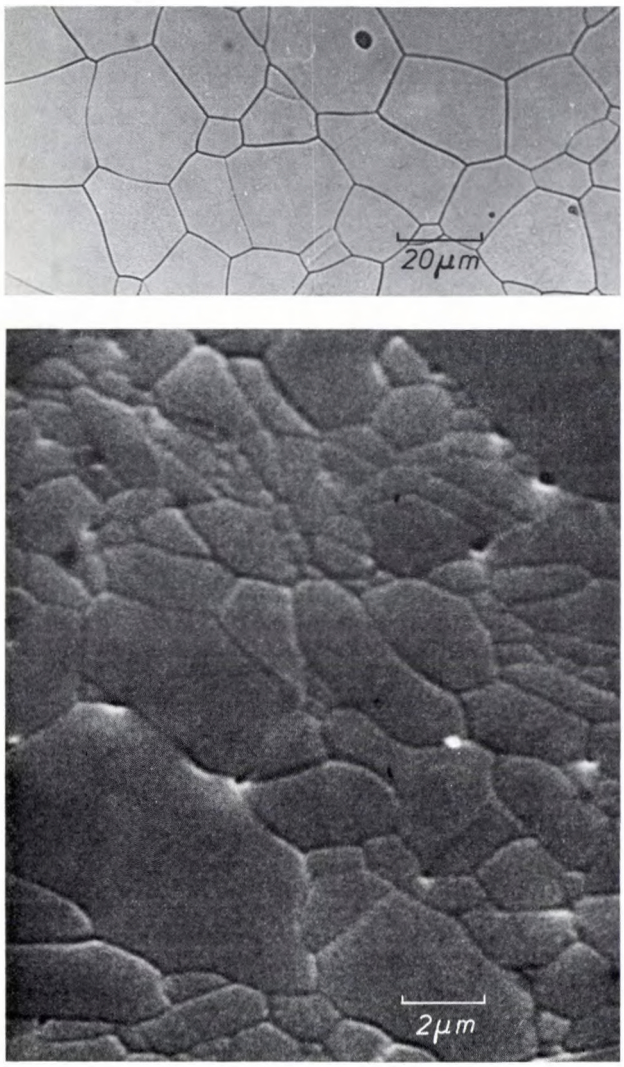


Fig. 5. The microstructure of sintered (*above*) and of hot-pressed (*below*) alumina. The sintered material contains a little magnesium oxide, which accounts for its homogeneous structure. The crystallites of the hot-pressed material are approximately a factor of 10 smaller.

view of the high purity of our starting material and the preparation process used, particles of a second phase may be disregarded. There are, however, pores in the material, and their number and size do of course depend very closely on the manner in which the material is made.

Material samples

As mentioned earlier, our measurements were carried out both on 'ordinary' sintered alumina samples and on samples obtained by continuous hot pressing. The starting material for the sintered samples was alumina powder with an addition of 0.03-0.2% of magnesium oxide. Sintering took about 10 hours and was carried out in a humid hydrogen atmosphere at a temperature of 1850 °C. No magnesium oxide was added to the hot-pressed material. The temperature was 1400 °C, the pressure 1200 atm and the time 30 minutes; the process took place in an oxygen atmosphere. *Fig. 5* shows the microstructure in the two cases. The sintered material consists of grains with an average diameter of about 25 μm. The pores are spherical (radius 0.5-1 μm) and they lie both inside the grains and at the grain boundaries. In the hot-pressed material it is noticeable that the structure is less homogeneous — no magnesium oxide was added — and grain sizes lie mainly between 1 and 2 μm. In this case the pores are very much smaller (radius about 0.1 μm), and most of them lie at the grain boundaries.

The hot-pressed material thus consists of much smaller grains than the sintered material. In spite of this, the light scattering by grain boundaries or the birefringence at these boundaries is apparently not very significant, for otherwise the material would not have such a narrow scattering profile (fig. 4). X-ray analyses and electron-microscope examinations have shown that in the hot-pressed alumina there is very little if any preferential orientation of the grains normal to or parallel to the pressing direction. If preferential orientation existed it could certainly have some effect on the light transmission.

The porosity of the material is exceptionally low in both cases; the volume fraction of the pores is no more than about 0.1% or less. Although this figure appears to be very satisfactory, it should be remembered that it is derived from a difference, which makes it relatively uncertain. In the calculations of light scattering the porosity was therefore generally taken as a free parameter and not as a known quantity.

Calculation of the light transmission

The theoretical calculations will evidently serve our purpose if measured results like those given in fig. 3 and fig. 4 can be described as the effect of light scattering by pores still present in the samples.

The in-line transmission of light trying to pass through an alumina layer with homogeneously distributed pores decreases exponentially with the thickness of the layer — provided that the layer is sufficiently thin and the porosity sufficiently low [7]. A relationship of this kind is also found theoretically when it is assumed that every pore which the light beam encounters in its passage through the layer has a scattering effect independent of the others, and also that a light ray is never scattered more than once:

$$I/I_0 = K \exp(-\gamma d).$$

In this equation d is the layer thickness, calculated in the direction of incidence of the light, and γ is the

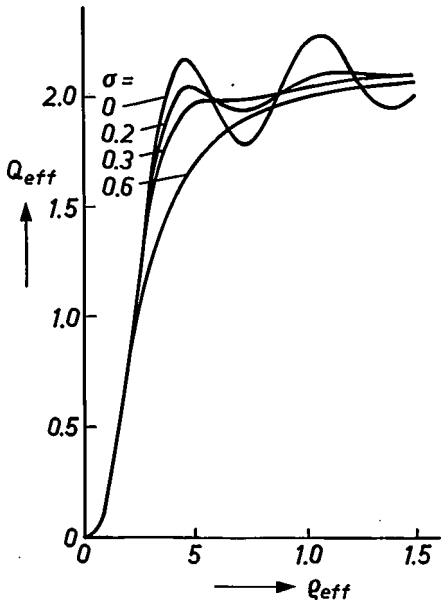


Fig. 6. The effective scattering efficiency Q_{eff}, calculated as a function of the quantity ϱ_{eff}, which is a measure of the effective pore radius r_{eff}. The parameter σ is the standard deviation of the assumed logarithmic Gaussian distribution of the pore radii. Experimental results for the distribution of the pore radii fit reasonably well for a σ of 0.2 to 0.3.

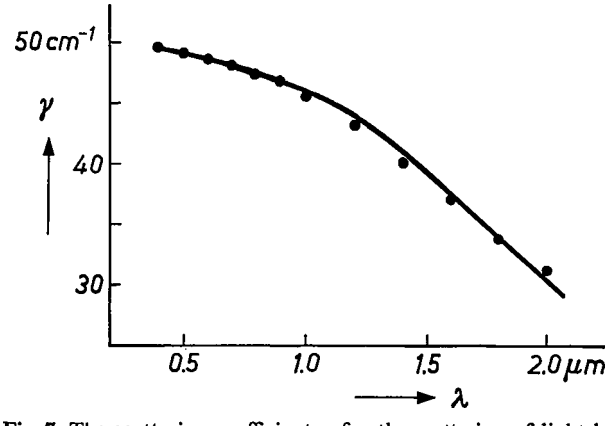


Fig. 7. The scattering coefficient γ for the scattering of light by pores in a thin layer of sintered alumina, plotted against the wavelength λ of the incident light. The points were obtained from transmission measurements. The curve is the calculated best fit to the experimental results; see Table I.

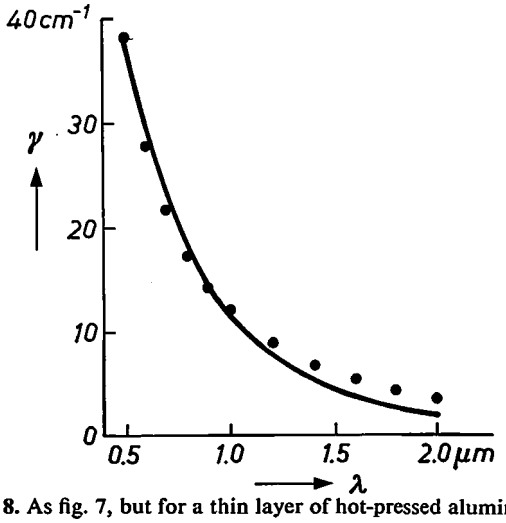


Fig. 8. As fig. 7, but for a thin layer of hot-pressed alumina.

scattering coefficient. The factor K takes account of the reflection loss, which will not be dealt with here.

The method of calculating the transmission amounts to the prediction of the scattering coefficient as a function of the known data for the pores. The pores are assumed to be spherical with radii that follow a logarithmic Gaussian distribution. 'Logarithmic' here means that it is not the pore radius itself but its logarithm that follows a Gaussian distribution [8]. It has been found from photomicrographs of polished material cross-sections that this logarithmic Gaussian distribution represents the reality fairly well in our samples.

The calculations include the volume fraction V_p of the pores, the 'effective pore radius' r_{eff} — i.e. the radius that identical pores should have to cause the same light scattering as in the logarithmic Gaussian distribution — and the standard deviation σ. The quantity σ is a measure of the width and also of the skewness of the distribution; because of this the skewness of the distribution increases with its width.

The scattering coefficient may be described as the product of the 'effective scattering efficiency' Q_{eff} and the total scattering surface $N\pi r_{eff}^2$. The quantity N is the number of pores per unit volume; the total scattering surface is calculated in the light path and per unit of illuminated surface. The scattering efficiency is calculated with a computer from Mie's equations [5]. The values found are given in *fig. 6* as a function of the quantity ϱ_{eff}, which is equal to $4\pi(n_{Al_2O_3} - n_{pore})r_{eff}/\lambda$, where λ is the wavelength of the light (in vacuum) and n is the refractive index. As can be seen, the standard deviation has a considerable influence on the form of the curve. Oscillations in the curve, like those for example for $\sigma = 0$, were never observed in our experiments, which indicates that any assumption of a uniform pore size would be a long way from the truth.

The agreement between the calculated results and the experimental results for in-line transmission can clearly be seen from *fig. 7* and *fig. 8*. The first figure relates to measurements for a sintered sample, the second for a hot-pressed sample. In both figures the scattering coefficient was determined from transmission measurements at twelve wavelengths. The curve shown is the best theoretical fit to the experimental results, and was obtained by using a least-squares optimization in the choice of the effective pore radius, the standard deviation of the pore radii and the volume fraction.

[8] The logarithmic Gaussian distribution applies exactly for quantities that are the result of a process involving steps where the probability of a change is proportional to the instantaneous value of the quantity. Many growth and breakdown processes belong to this category. See for example J. Aitchison and J. A. C. Brown, The lognormal distribution, Cambridge Univ. Press, 1957.

Table I. The values that give the best fit between theory and experiment for the effective pore radius r_{eff}, the standard deviation σ of the pore radii and the fraction V_p of the volume occupied by all pores, for hot-pressed material ('transparent') and sintered material ('translucent'). The quantity r_m is the value of the pore radius for which there is a maximum in the probability distribution.

	Hot pressed	Sintered
r_{eff}	0.15 μm	0.59 μm
r_m	0.11 μm	0.34 μm
σ	0.3	0.4
V_p	4.9×10^{-4}	1.9×10^{-3}

The values thus determined are listed in *Table I*. In view of the relatively large number of wavelengths used in the measurements, including wavelengths outside the visible region, the procedure used for the fit may be regarded as reasonably reliable. As can be seen from the Table, the pores in the hot-pressed material — the 'transparent' material — are much smaller than in the sintered material. The values found agree well with those determined from plots like those of fig. 5.

No calculations have yet been carried out on the exact shape of the scattering profile (fig. 4). It may be assumed on experimental grounds, however, that the profile will become narrower as the number of pores and the pore radii decrease and the radii come more into the region of Rayleigh scattering.

Summary. The relationship between the microstructure of polycrystalline alumina (Al_2O_3) and its light transmission has been investigated. The transmission loss measured in thin layers is largely attributable to Mie scattering from the pores in the material; the reflection losses are about 15%. The effect of absorption and birefringence is negligible. The layers investigated were made by conventional sintering (MgO added) or by continuous hot pressing. The porosity is about 0.1%. The radius of the pores left in the sintered alumina is 0.5 to 1 μm, and in the hot-pressed material about 0.1 μm. Layers of the first type are diffusely translucent. The other type of material has a much narrower scattering profile, and is therefore more transparent. The material is used for the envelopes of high-pressure sodium lamps.

Recent 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*
Mullard Research Laboratories, Redhill, Surrey, England *M*
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*
MBLE Laboratoire de Recherches, 2 avenue Van Becelaere, 1170 Brussels (Boitsfort), Belgium *B*
Philips Laboratories, 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A. (by contract with the North American Philips Corp.) *N*

Reprints of most of these publications will be available in the near future. Requests for reprints should be addressed to the respective laboratories (see the code letter) or to Philips Research Laboratories, Eindhoven, The Netherlands.

E. A. Aagaard, J. W. Coenders & E. C. Dijkmans: P-N-P-N interface-to-interface switching for telephony and broad-band communication.
IEEE Trans. **COM-22**, 1250-1263, 1974 (No. 9). *E*

M. Audier & J. P. Boutot: Multidetector PM with microchannel plate.
Philips Res. Repts. **30**, 226-240, 1975 (No. 4). *L*

J. Basterfield, J. M. Shannon & A. Gill: The nature of barrier height variations in alloyed Al-Si Schottky barrier diodes.
Solid-State Electronics **18**, 290-291, 1975 (No. 3). *M*

D. Bäuerle: Vibrational spectra of electron and hydrogen centers in ionic crystals.
Springer Tracts in Modern Physics **68**, 77-160, 1973. *A*

H. C. J. van Beukering & H. Fokker: De stirlingmotor. Een alternatief voor de benzinemotor.
Natuur en Techniek **42**, 400-409, 1974 (No. 7). *E*

J. Bloem & B. Bosman: Etching of Al_2O_3 in an HCl-H_2 ambient at 1200-1500 K.
Philips Res. Repts. **30**, 206-217, 1975 (No. 4). *E*

J. H. den Boef & J. C. M. Henning: Design of a strain modulation ESR spectrometer.
Proc. 18th Ampere Congress, Nottingham 1974, Vol. I, pp. 197-198; 1975. *E*

M. Bouckaert, A. Pirotte & M. Snelling: Efficient parsing algorithms for general context-free parsers.
Information Sci. **8**, 1-26, 1975 (No. 1). *B*

M. Brouha & A. G. Rijnbeek: Curie temperature determination from permeability measurements up to 150 kbar. Results for nickel.
High Temp. - high Press. **6**, 519-524, 1974 (No. 5). *E*

E. Bruninx: The coprecipitation of Zn, Cd and Hg with ferric hydroxide.
Philips Res. Repts. **30**, 177-191, 1975 (No. 4). *E*

E. Bruninx & H. J. Prins: The accurate dead time measurement of single channel gamma ray counters by means of the proportional source method.
Int. J. appl. Rad. Isot. **25**, 483-492, 1974 (No. 11/12). *E*

K. H. J. Buschow: Magnetic properties of CsCl-type rare-earth cadmium compounds.
J. chem. Phys. **61**, 4666-4670, 1974 (No. 11). *E*

K. H. J. Buschow, M. Brouha & A. G. Rijnbeek: Thermomagnetic history effects in $LaCo_{5x}Ni_{5-5x}$ and $ThCo_{5x}Ni_{5-5x}$ compounds.
Solid State Comm. **16**, 31-34, 1975 (No. 1). *E*

B. L. Cardozo (Institute for Perception Research, Eindhoven): Some notes on frequency discrimination and masking.
Acustica **31**, 330-336, 1974 (No. 6).

F. M. A. Carpay: On the eutectoid reaction in Cu-Al and Fe-C.
Metallurg. Trans. **5**, 2614-2615, 1974 (No. 12). *E*

H. T. van Dam & J. J. Ponjée: Electrochemically generated colored films of insoluble viologen radical compounds.
J. Electrochem. Soc. **121**, 1555-1558, 1974 (No. 12). *E*

H. Dammann & E. Klotz: Generation of faultless multi-pinhole masks by means of spatial filtering.
Optics Comm. **13**, 268-272, 1975 (No. 3). *H*

N. H. Dekkers & H. de Lang: Differential phase contrast in a STEM.
Optik **41**, 452-456, 1974 (No. 4). *E*

C. Z. van Doorn: Transient behaviour of a twisted nematic liquid-crystal layer in an electric field.
J. Physique **36**, C1/261-263, 1975 (Colloque C1). *E*

W. F. Druyvesteyn, F. A. Kuijpers, E. de Niet & A. W. M. van den Enden: An all-permalloy bubble memory with control functions based on drive-field operation.
IEEE Trans. **MAG-10**, 757-760, 1974 (No. 3). *E*

O. Elgersma & G. Voorn: Deconvolution of absorption spectra at room temperature of chloroplast fragments.
Proc. 3rd Int. Congress on Photosynthesis, Rehovot, Israel 1974, pp. 1943-1949. *E*

P. van Engelen: ENDOR of GaP:Mn.
Proc. 18th Ampere Congress, Nottingham 1974, Vol. II, pp. 445-446; 1975. *E*

G. Engelsma: On the mechanism of the changes in phenylalanine ammonia-lyase activity induced by ultraviolet and blue light in gherkin hypocotyls.
Plant Physiol. **54**, 702-705, 1974 (No. 5). *E*

J. van Esdonk & J. F. M. Janssen: Joining a sputtering target and a backing plate.
Research/Development **26**, No. 1, pp. 41, 43 & 44, Jan. 1975. *E*

B. Fischer, W. J. **Buckel** (both with Max-Planck-Institut für Festkörperforschung, Stuttgart) **& D. Bäuerle:** Surface polaritons in tetragonal $BaTiO_3$ and $PbTiO_3$.
Solid State Comm. **15**, 1801-1804, 1974 (No. 11/12). *A*

A. J. Fox: Nonlinear longitudinal KTN modulator.
Appl. Optics **14**, 343-352, 1975 (No. 2). *M*

F. W. Gorter, J. A. L. Potgiesser & D. L. A. Tjaden: Magnetoresistive reading of information.
IEEE Trans. **MAG-10**, 899-902, 1974 (No. 3). *E*

M. Goscianski & L. Léger: Electrohydrodynamic instabilities above a nematic to smectic *A* (or *C*) transition.
J. Physique **36**, C1/231-235, 1975 (Colloque C1). *L*

G. Groh & M. Kock: Holographic archival of X-ray images.
Medicamundi **19**, 43-44, 1974 (No. 2). *H*

J. C. M. Henning & J. H. den Boef: Strain modulated electron spin resonance.
Proc. 18th Ampere Congress, Nottingham 1974, Vol. I, pp. 117-118; 1975. *E*

S. Herman: The design of television color rendition.
J. SMPTE **84**, 267-273, 1975 (No. 4). *N*

D. d'Humières & L. Léger: Critical behaviour above a nematic to smectic *A* phase transition.
J. Physique **36**, C1/113-116, 1975 (Colloque C1). *L*

L. Jacomme: A model for ray propagation in a multimode graded-index fibre.
Optics Comm. **14**, 134-138, 1975 (No. 1). *L*

B. A. Joyce: Growth of single crystals of GaAs in bulk and thin film form.
Crystal growth, editor B. R. Pamplin, publ. Pergamon, Oxford 1975, pp. 157-184. *M*

E. T. Keve & K. L. Bye: Phase identification and domain structure in PLZT ceramics.
J. appl. Phys. **46**, 810-818, 1975 (No. 2). *M*

A. J. R. de Kock, P. J. Roksnoer & P. G. T. Boonen: Formation and elimination of growth striations in dislocation-free silicon crystals.
J. Crystal Growth **28**, 125-137, 1975 (No. 1). *E*

H. Koelmans: Metallization corrosion in silicon devices by moisture-induced electrolysis.
IEEE 1974 Reliability Physics Symp., Las Vegas (12th ann. Proc.), pp. 168-171. *E*

H. K. Kuiken: The thick free-convective boundary-layer along a semi-infinite isothermal vertical cylinder.
Z. angew. Math. Physik **25**, 497-514, 1974 (No. 4). *E*

W. Kwestroo, H. A. M. van Hal & C. Langereis: Compounds in the system BaO-Sc_2O_3.
Mat. Res. Bull. **9**, 1623-1629, 1974 (No. 12). *E*

W. Kwestroo, H. A. M. van Hal & C. Langereis: Compounds in the system BaO-Y_2O_3.
Mat. Res. Bull. **9**, 1631-1637, 1974 (No. 12). *E*

P. K. Larsen & R. Metselaar: Defects and the electronic properties of $Y_3Fe_5O_{12}$.
J. solid State Chem. **12**, 253-258, 1975 (No. 3/4). *E*

W. Lems, H. Kinkartz & W. Lechner: The influence of facet formation on the life of incandescent lamps.
Philips Res. Repts. **30**, 218-225, 1975 (No. 4). *A*

A. Mank: Solid pentomino multiplications.
J. recreat. Math. **7**, 279-282, 1974 (No. 4). *E*

G. Martin, P. Bach, J. C. Tranchart & E. Fabre: Optical quenching in CdTe detectors.
IEEE Trans. **NS-22**, 226-228, 1975 (No. 1). *L*

R. J. Meijer & C. L. Spigt: The potential of the Philips Stirling engine for pollution reduction and energy conservation.
2nd Symp. on Low pollution power systems development, Düsseldorf 1974, 12 pp. *E*

R. Memming: Photochemical processes in monomolecular dye layers on SnO_2.
Faraday Disc. Chem. Soc. No. 58, 261-270, 1974. *H*

A. Mircea-Roussel, L. Léger, F. Rondelez & W. H. de Jeu: Measurements of transport properties in the nematic and smectic phases of various compounds.
J. Physique **36**, C1/93-97, 1975 (Colloque C1). *L, E*

A. E. Morgan: Liquid phase epitaxial growth kinetics of magnetic garnet films grown by isothermal vertical dipping from unstirred fluxed melts.
J. Crystal Growth **27**, 266-273, 1974. *H*

P. P. Mulder (Philips Lighting Division, Eindhoven): Practische aspecten van de Kr^{85}-lekdetector.
Ned. T. Vacuümtechniek **13**, 8-13, 1975 (No. 1).

K. H. Nicholas, R. A. Ford, P. J. Daniel, C. W. Sullivan (Associated Semiconductor Manufacturers, Southampton), **P. Sant** (Ass. Semicond. Manuf., Southampton), **C. Bull** (University of Oxford) **& G. R. Booker** (Univ. of Oxford): Reduced gain of ion-implanted transistors.
Appl. Phys. Letters **26**, 320-322, 1975 (No. 6). *M*

W. J. Oosterkamp: New concepts and progress in instrumentation for cine- and video-radiology.
Medicamundi **19**, 79-84, 1974 (No. 3). *E*

W. J. Oosterkamp & A. P. M. van 't Hof: Colour radiography and subtraction.
Medicamundi **19**, 116-118, 1974 (No. 3). *E*

A. van Oostrom: Investigation of vacuum breakdown using Auger spectroscopy.
Proc. 2nd Int. Conf. on Solid surfaces, Kyoto 1974 (Jap. J. appl. Phys. Suppl. 2, Part 2, 1974), pp. 795-801. *E*

P. Piret: On a class of alternating cyclic convolutional codes.
IEEE Trans. **IT-21**, 64-69, 1975 (No. 1). *B*

H. Rau: *p-v-T* relations for HCl.
High Temp. - high Press. **6**, 671-673, 1974 (No. 6). *A*

M. H. J. van Rijswick: Adaptive program for high precision off-line processing of chromatograms.
Chromatographia **7**, 491-501, 1974 (No. 9). *E*

J. M. Robertson, W. Tolksdorf & H. D. Jonker: Growth mechanisms and composition in the LPE process for bubble domain materials.
J. Crystal Growth **27**, 241-248, 1974. *E, H*

P. Schagen: Image converters and intensifiers.
J. Physics E **8**, 153-160, 1975 (No. 3). *M*

A. Schmitz & A. Slob: The effect of isolation regions on the current gain of inverse npn transistors used in Integrated Injection Logic (I^2L).
1974 IEDM Tech. Digest, pp. 508-510. *E*

J. P. Scott: Photocathodes for use in an electron image projector.
J. appl. Phys. **46**, 661-664, 1975 (No. 2). *M*

P. J. Severin: Formal comparison of correction formulae for spreading resistance measurements on layered structures.
Nat. Bur. Stand. special Publn. 400-10, pp. 27-44, 1974. *E*

P. J. Severin: Angle-bevelling silicon epitaxial layers, technique and evaluation.
Nat. Bur. Stand. special Publn. 400-10, pp. 99-108, 1974. *E*

P. J. Severin & H. Bulle: Four-point probe measurements on n-type silicon with mercury probes.
J. Electrochem. Soc. **122**, 133-137, 1975 (No. 1). *E*

P. J. Severin & H. Bulle: Spreading resistance measurements on n-type silicon using mercury probes.
J. Electrochem. Soc. **122**, 137-142, 1975 (No. 1). *E*

M. Sintzoff: A brief review of ALGOL 68.
ALGOL Bull. No. 37, 54-62, 1974. *B*

J. L. Sommerdijk & J. M. P. J. Verstegen (Philips Lighting Division, Eindhoven): Concentration dependence of the Ce^{3+} and Tb^{3+} luminescence of $Ce_{1-x}Tb_xMgAl_{11}O_{19}$.
J. Luminescence **9**, 415-419, 1974 (No. 5). *E*

M. J. Sparnaay: On the effect upon substrate surface properties of very small amounts of an adsorbate.
Proc. 2nd Int. Conf. on Solid surfaces, Kyoto 1974 (Jap. J. appl. Phys. Suppl. 2, Part 2, 1974), pp. 299-302. *E*

M. J. Sparnaay: Ellipsometry.
Dechema-Monographien No. 78, 247-270, 1974. *E*

W. T. Stacy & M. M. Janssen: X-ray Pendellösung in garnet epitaxial layers.
J. Crystal Growth **27**, 282-286, 1974. *E*

G. H. A. M. van der Steen & E. Papanikolau (Philips Glass Development Centre, Eindhoven): Introduction and removal of hydroxyl groups in vitreous silica, Part II. Chemical and physical solubility of hydrogen in vitreous silica.
Philips Res. Repts. **30**, 192-205, 1975 (No. 4).

G. E. Thomas & E. E. de Kluizenaar: A chemical effect on light emission from ion-bombarded copper and aluminium surfaces.
Int. J. Mass Spectrom. Ion Phys. **15**, 165-179, 1974 (No. 2). *E*

N. C. de Troye: Integrated injection logic — present and future.
IEEE J. SC-**9**, 206-211, 1974 (No. 5). *E*

J. van der Veen: The influence of terminal substituents upon the nematic-isotropic transition temperature.
J. Physique **36**, C1/375-377, 1975 (Colloque C1). *E*

J. van der Veen & W. H. de Jeu: Liquid crystalline bridge substituted Schiff bases.
Mol. Cryst. liq. Cryst. **27**, 251-257, 1974 (No. 3/4). *E*

J. M. P. J. Verstegen (Philips Lighting Division, Eindhoven), **J. L. Sommerdijk & A. Bril:** Line emission of $SrAl_{12}O_{19}:Eu^{2+}$.
J. Luminescence **9**, 420-423, 1974 (No. 5). *E*

J. M. P. J. Verstegen (Philips Lighting Division, Eindhoven) **& A. L. N. Stevels:** The relation between crystal structure and luminescence in β-alumina and magnetoplumbite phases.
J. Luminescence **9**, 406-414, 1974 (No. 5). *E*

H. J. Vink: Scientific understanding of the manufacture and properties of solid oxidic industrial materials.
Defects and transport in oxides, editors M. S. Seltzer & R. I. Jaffee, publ. Plenum Press, New York 1974, pp. 127-138. *E*

W. F. van der Weg (Philips Research Labs., Amsterdam Division) **& E. Lugujjo** (California Institute of Technology, Pasadena): Optical line and broad-band emission from ion-bombarded targets.
Atomic collisions in solids, Vol. 2, editors S. Datz, B. R. Appleton & C. D. Moak, publ. Plenum Press, New York 1974, pp. 511-522.

H. W. Werner: Secondary ion mass spectrometry and its application in thin film and surface layer research.
Proc. 6th Int. Vacuum Congress, Kyoto 1974 (Jap. J. appl. Phys. Suppl. 2, Part 1, 1974), pp. 367-374. *E*

H. W. Werner: Theoretical and experimental aspects of secondary ion mass spectrometry.
Vacuum **24**, 493-504, 1974 (No. 10). *E*

J. S. C. Wessels & M. T. Borchert: Studies on subchloroplast particles. Similarity of grana and stroma photosystem I and the protein composition of photosystem I and photosystem II particles.
Proc. 3rd Int. Congress on Photosynthesis, Rehovot, Israel 1974, pp. 473-484. *E*

C. H. Weysenfeld: Small dye laser in a semiunstable resonator pumped by an argon-jet guided spark.
Appl. Optics **13**, 2816-2822, 1974 (No. 12). *E*

M. V. Whelan: Minority-carrier injection and transient response of a MOS capacitor.
Philips Res. Repts. **30**, 262-264, 1975 (No. 4). *E*

M. V. Whelan: Silicon-surface space-charge strong-inversion capacitance and minority-carrier injection.
Philips Res. Repts. **30**, 265-268, 1975 (No. 4). *E*

M. V. Whelan & L. A. Daverveld: Novel semiconductor switch with controllable delay and rapid turn-on.
Philips Res. Repts. **30**, 256-261, 1975 (No. 4). *E*

R. F. Wielinga: Constrained interpolation using Bézier curves as a new tool in computer aided geometric design.
Computer aided geometric design (Proc. Conf. Salt Lake City 1974), editors R. E. Barnhill & R. F. Riesenfeld, publ. Academic Press, New York 1974, pp. 153-172. *E*

R. R. Wilson (NKF Kabel B.V., Delft) **& G. C. Groenendaal:** Calculation of the temperature dependence of the primary and secondary parameters of coaxial cable.
Philips Res. Repts. **30**, 241-255, 1975 (No. 4). *E*

H. J. de Wit & C. Crevecoeur: The dielectric breakdown of anodic aluminium oxide.
Physics Letters **50A**, 365-366, 1974 (No. 5). *E*

Contents of Philips Telecommunication Review **33**, No. 3, 1975:

M. Göhl & H. P. van Roosmalen: Electronic Private Automatic Branch Exchange EBX 100 (pp. 105-111).
R. T. van der Schaaf: EBX 8000, a stored program controlled PABX system (pp. 113-124).
R.A.D. Maltha & H.J. Goebertus: PRX 205 large scale introduction in the Dutch telephone network (pp. 125-137).
A. J. M. Dingjan & C. Ziekman: New designs for digital transmission equipment (pp. 138-143).
J. van der Vegte: Design-400 to house analog and digital transmission equipment (pp. 144-149).
A. van Dedem & H. Rebel: 8TR 401, a new generation of channel and group translating equipment (pp.150-158).
R. Azimullah: Semi-automatic ship-to-shore telex communication (pp. 159-162).
J. K. A. Poppe: Philips Negistors (pp. 163-167).
P. M. van Daal: Concentrated HF radio monitoring (pp. 168-172).

Contents of Electronic Applications Bulletin **33**, No. 2, 1975:

L. van de Meeberg: Analysis and measurement of the error-correcting properties of the Viterbi algorithm (pp. 45-58).
W. Th. Hetterscheid: Turn-on and turn-off behaviour of high-voltage switching transistors (pp. 59-72).
P. G. J. Barten & J. Kaashoek: Large-screen colour television with intrinsically convergent 110° deflection (pp. 75-89).
Circulator selection nomogram (pp. 90-91).

Contents of Mullard Technical Communications **13**, No. 128, 1975:

D. J. Beakhust & M. C. Gander: Complementary push-pull video amplifiers for television receivers (pp. 294-301).
L. Hampson: Fuse protection of semiconductor diodes, thyristors, and triacs (pp. 302-319).
Gapped ferroxcube E-cores for switched-mode power supplies (p. 320).
L. E. Jansson: Power-handling capability of ferrite transformers and chokes for switched-mode power supplies (pp. 321-352).

PHILIPS TECHNICAL REVIEW

VOLUME 36, 1976, No. 3

Electrostatic printing

U. Rothgordt

Writing and drawing have been used by man for thousands of years as methods for the storage and dissemination of information. The art of making printed books, allowing even more people to benefit from the information they contain, is already centuries old. Today it sometimes seems as though we are in danger of disappearing under an avalanche of printed paper. Let us only hope that new printing techniques will lead, not to an even greater avalanche of paper, but to a more sensibly directed, more efficient distribution of the information. In spite of the increasing importance of optical and acoustic means of communication, there is still a continuing need for methods that will provide information in 'black and white'. The design of printing equipment for this purpose that is also compatible with the requirements set by the use of computers for information processing, or perhaps by as yet undeveloped machines that will produce printed information at a terminal, is a challenge to industrial research and development. The results of such investigations, described in the article below, are machines that are eminently suitable for the rapid production of text on paper where not many copies are generally required.

Introduction

The electrostatic printing process is a non-mechanical process, which in its simplest form consists of two stages. In the first stage, a charge pattern is produced on a suitable surface by an electric field. In the second, this charge pattern is made visible and fixed by a suitable dye powder. It will become clear in the course of this article why the process is called 'electrostatic': it certainly does not mean that the actual process is 'static'. The description of 'electrostatic printing' is often applied incorrectly, for example to techniques in which a colour change in the paper to be printed is produced by passing a current through an electrosensitive paper. Here we shall confine ourselves to the two-stage process just mentioned.

Electrostatic printing is also sometimes confused with the process of electrophotography now widely used in graphic reproduction. In that process a uniform electric charge is applied to a photoconductive layer, which is then exposed to an image of the pattern to be copied. This produces local changes in the state of charge. The charge pattern is then developed, again by using a dye powder.

Electrostatic printers are applied as printers or plotters in data processing, and also in measurement techniques for recording measured data, e.g. in the form of plotted curves. The production of copies at a remote terminal is another interesting potential application. The coded information for the reproduction of a document is transmitted, e.g. over a telephone line, and the document is then reproduced in virtually the original form.

In all these applications mechanical printing methods have so far been the most widely used [1]; the non-mechanical methods that have been used include thermography, 'Ink jet', and the one mentioned earlier with electrosensitive paper. The high print rates that can be achieved with electrostatic printing (up to four pages of A4 paper per second) are of great interest in a number of these applications. In computer output devices high speeds are desirable, so that processing

Dr U. Rothgordt is with Philips GmbH Forschungslaboratorium Hamburg, Hamburg, West Germany.

[1] See for example J. Borne, Philips tech. Rev. **29**, 205, 1968.

times are not too greatly affected by the output rate. In recording devices for measurement applications the highest frequency that can be recorded is determined by the maximum paper speed. Another advantage of electrostatic printing is that it is a non-impact process, and therefore quiet in operation; the equipment will also be less subject to wear, which means that there will be few breakdowns.

Even within the limited definition of electrostatic printing used here, there are still many ways in which the basic principles can be put into practice. The charge can be supplied by a gas discharge from a thin stylus electrode, or by a small glow discharge. It is also possible to use very thin electrodes that can be moved rapidly over a dielectric surface, producing a charge pattern on it. Starting from these basic ideas we have been able to develop equipment for very widely differing applications. *Fig. 1* shows an example of such equipment: a 'visible speech recorder'. This is an instrument that will record acoustic signals in real time; we shall return to its operation at the end of the article.

In the following sections we shall look more closely at the various ways in which a charge pattern can be produced, and also at the requirements for the dielectric substrate. Besides these methods for applying a charge pattern, the methods for development are also of obvious importance. Details of the methods that we have most frequently used will also be discussed.

However, before embarking on a more detailed description of the printing process, it is perhaps of interest to briefly mention a few points from the history of the process [2].

Although the existence of static electricity has been known for longer than that of electric currents, it was not until the twenties of this century that P. Selényi [3] made the first attempt to make use of electrostatic phenomena for a printing process. He started by using a cathode-ray tube whose screen was dusted with lycopodium powder on the outside after the beam had written a charge pattern on the inside. This produced a copy of the charge pattern. Further stages in the development were the application of a charge pattern to a dielectric layer inside a cathode-ray tube, a cathode-ray tube with an electron-transmitting window (a Lenard window) with which a charge pattern could be applied to a dielectric substrate outside the tube, and finally an ion beam that could be deflected in air. With the ion-beam method Selényi was even able to record television pictures. However, none of his instruments has ever been of any practical significance.



Fig. 1. An electrostatic printer used for the real-time presentation of acoustic spectra. The printer has been opened for loading with paper.

Technical applications of 'electrostatography' only appeared after 1950, after the promising development of C.F. Carlson's electrophotography had provided the incentive. Work was done on a cathode-ray tube in which the Lenard window used by Selényi was replaced by a screen with a large number of copper wires, all

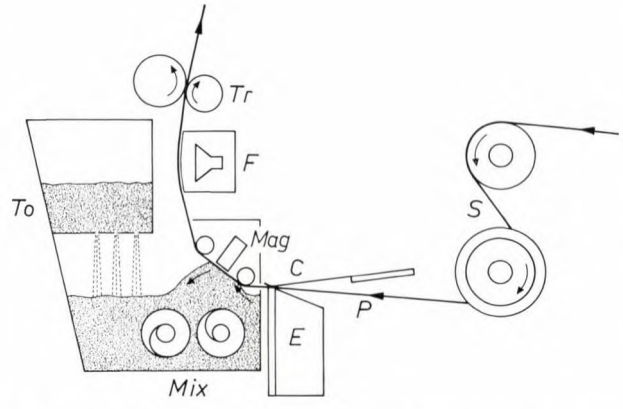


Fig. 2. Diagrammatic cross-section of an electrostatic printer. The paper to be printed, which is coated with a dielectric layer, is drawn through the machine by the transport roller *Tr*; the tension and synchronization rollers *S* ensure that the paper moves in the correct steps. The paper is charged by the 'comb' of pin electrodes *E* and the counter electrode *C*. The paper then travels to the developer unit. Here toner and iron powder are continuously mixed together in the mixing chamber *Mix*. The mixture is drawn against the paper by the magnet *Mag*. The toner particles that have become attached to the charged areas of the paper are bonded to the paper by heating briefly in the fixer unit *F*; this fixes the image. The developer powder is continuously replenished with toner from the supply *To* to make up for the toner that has been used.

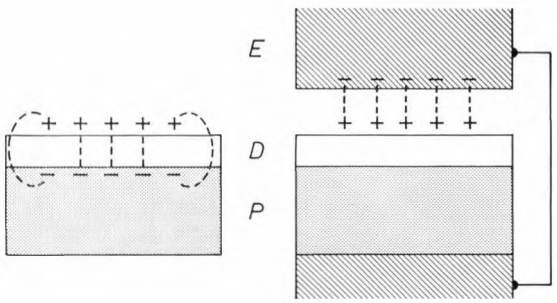


Fig. 3. Schematic diagram of the lines of force around a charged location on the substrate. *D* dielectric layer. *P* paper. *Left:* with no external electrode. The lines of force are mainly *inside* the dielectric layer and only emerge from the substrate at the edge of the charged area. Toner particles will in the main only adhere at such edges. *Right:* with an external developer electrode *E*. The lines of force now emerge from the substrate everywhere, giving uniform developing.

insulated from each other and at right angles to the screen [4]. This arrangement gave a fairly high resolution in the charge image formed on a dielectric surface pressed against the outside of the screen. In another case a matrix of electrode pairs was used [5]. A gas discharge was produced between certain pairs, and some of the charge carriers from the discharge were deposited on a dielectric substrate. Special precautions were necessary to prevent too rapid erosion of the electrodes. This was the experience available to us in the sixties when we started an investigation to find new methods of electrostatic printing; the results of this work are discussed in this article [6].

The principle of an electrostatic printer

An explanation of the printing process [7] will be given with the aid of a cross-section of an electrostatic printer (*fig. 2*). The paper to be printed passes through a tensioning device, and then arrives at the printing section, consisting of a printing head and a counter-electrode array. The printing head consists of a row of closely spaced pin electrodes, which can apply to the paper points of charge corresponding to the desired image. The paper is coated with a special dielectric layer.

After charging, the paper travels to the developer unit, where there is a mixture of coarse iron powder and a fine dye powder, the 'toner'. The toner particles acquire a charge through friction against the iron particles, and therefore adhere to them. A magnet placed behind the paper draws the powder mixture against the paper, bringing the charged toner particles into contact with the charged paper surface. The toner particles, which carry a charge of opposite sign to that applied to the paper, then adhere to the paper at the areas covered by the charge pattern, so that the charge distribution is made visible. In this process the iron powder has a two-fold function. Unlike the fine toner powder (particle size between 5 and 30 μm), the much coarser iron powder (particle size 100-400 μm) flows very freely, which ensures a uniform distribution of the toner particles over the charge image to be developed. In addition to this the iron powder brings the electrostatic field outside the dielectric layer, because of its electrical conductivity and the good contact between the particles (*fig. 3*). The field can thus attract the toner particles. The field-strength outside the dielectric layer increases as the external electrode (the iron powder) is brought closer to the layer. The diameter of the toner particles should not therefore be much larger than the thickness of the dielectric layer.

The third stage in the process is that of fixing, which permanently attaches to the paper the dye particles that were originally held in position by electrostatic and Van der Waals forces. In the dry development

[2] See for example R. M. Schaffert, Electrophotography, Focal Press, London 1965, and J. H. Dessauer and H. E. Clark (eds), Xerography and related processes, Focal Press, London 1965.

[3] P. Selényi, Z. techn. Physik **16**, 607, 1935.

[4] R. W. Crews and P. Rice, IRE Trans. **ED-8**, 406, 1961.

[5] H. Epstein, Proc. 1955 Western Joint Computer Conf., Los Angeles 1955, p. 116.
R. S. Howell, West German Patent Application No. 1 136 519, 13th Sept. 1962.

[6] Many colleagues have contributed over the years to the investigations described here. I should particularly like to name Dr R. Lorenz, whose share in the investigations was considerable, especially in the earlier stages.

[7] C. P. Hodum, U. Rothgordt and J. Schramm, in: Philips, Unsere Forschung in Deutschland, part II, p. 182, 1972.
H. F. Frohbach, IEEE Trans. **ED-19**, 579, 1972.

process described above fixing is usually accomplished by heating: the toner particles melt and adhere to the paper. The most practical arrangement is to heat the back of the paper, e.g. by passing it over a heated plate. After fixing the toner cannot be removed.

It may be desirable to generate the charge image, not on the actual paper to be distributed, but on a transfer surface that can be re-used. After development the toner image can be transferred by an electrostatic field to ordinary paper, and then fixed. The electrostatic field on the transfer surface can then be developed again, and another copy made on paper, or the surface may be cleaned electrically and a new charge pattern generated.

The stages described here are found in one form or another in all the variations of the electrostatic printing process to be described.

Charging the substrate

Before starting upon a description of the various methods of producing a charge image, it will be useful to look more closely at the physical effects that occur with a single electrode and a dielectric substrate. The substrate can be paper, coated with a thin dielectric layer, it can be a plastic (materials such as 'Mylar', 'Hostaphan', polyethylene and polycarbonate are suitable) or a ceramic layer on a rigid base, as used for a transfer surface for an image to be transferred to ordinary paper. Quantities of importance in all these substrates are the capacitance per unit area of the dielectric layer, the time constant of this capacitance and the conductivity of the paper. The capacitance determines the amount of charge per unit area that might reasonably be applied. The normal development processes require a charge density of a few times 10^{-4} C/m^2. For charging to 100 V the capacitance should be at least several microfarads per square metre.

The time constant should be at least high enough to give no noticeable loss of charge between charging and development; a value of several minutes is therefore desirable. This requirement sets no problems, since substrates are available that will hold a charge image for months.

Fig. 4 gives a diagrammatic representation of a single electrode with counter electrode, on opposite sides of the paper with its dielectric layer. The paper is lightly pressed against the counter electrode (10^4 to 10^5 N/m^2) to keep the contact impedance sufficiently low. The charging electrode is held about 10 μm from the surface of the dielectric layer.

If a large number of electrodes are embedded in a block of dielectric material, the correct spacing from the paper can be obtained by blowing air between the electrode block and the paper; the roughness of the paper surface may itself be sufficient to ensure the correct spacing. The air gap governs the whole process of charging, since charge can only arrive at the substrate when charge carriers are able to cross the gap. This can happen with a self-sustained gas discharge.

It might appear that there would be some danger of the electrodes burning away in such a discharge. In fact, however, the discharge extinguishes as soon as the potential of the dielectric layer has reached a certain value; the charge involved here is relatively small, so that the energy dissipation is also small, and the electrodes do not burn away.

The breakdown of a gas discharge is most easily described by the Paschen curve. This gives the relation between the electrode spacing, the gas pressure and the breakdown voltage of a self-sustained discharge for the case of two parallel planar electrodes. We shall not consider the physical basis of the Paschen curve further here [8]. *Fig. 5* gives the Paschen curve for air at a pressure of 1 atm and shows that the breakdown voltage then has a minimum at an electrode spacing of 8 μm. At this spacing the substrate can be fully charged at a voltage as low as 380 V. The total voltage that must be applied between a pin electrode and the counter electrode is however higher than the breakdown voltage of the discharge, because of the voltage drop across the paper and the contact impedance between paper and counter electrode. *Fig. 6* shows an equivalent circuit for the arrangement. It should be remembered here that the resistance R_d of the dielectric layer is very high, and that the capacitance C_d of this layer and C_g of the air gap are larger than the capacitance C_p of the paper because of the greater thickness of the paper. It is necessary to keep the paper resistance relatively low, to ensure that the time constant of the paper ($\tau_p = R_p C_p$) is low enough for short voltage pulses to break down

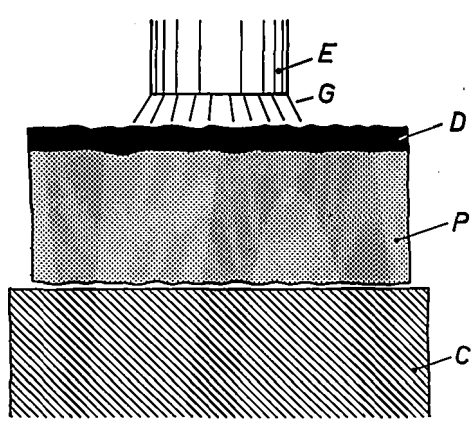


Fig. 4. Charging by a pin electrode. *E* electrode. *G* gas discharge. *D* dielectric layer. *P* paper; the paper is made to conduct by the addition of hygroscopic salts. *C* counter electrode.

the discharge. At a pulse length of 10 μs the time constant τ_p must be no greater than 1 μs. Since R_p is the paper resistance per unit area, it is equal to $\varrho_p d$, where ϱ_p is the resistivity of the paper and d is its thickness. Similarly, the capacitance per unit area C_p is equal to ε_p/d, where ε_p is the permittivity of the paper, so that $\tau_p = \varrho_p \varepsilon_p$. The relative permittivity of the paper is about 2.5, which means that ϱ_p should not be greater than about 4×10^6 Ω cm.

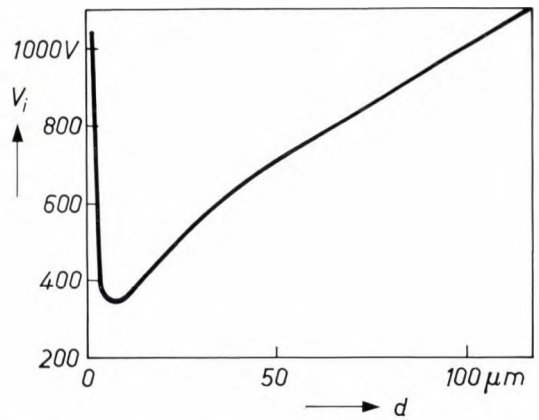


Fig. 5. The breakdown voltage V_i for a gas discharge in air at a pressure of 1 atm as a function of the width d of the air gap (the Paschen curve).

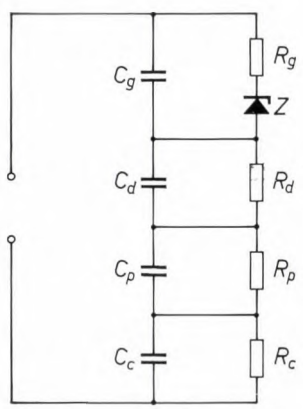


Fig. 6. Equivalent circuit for the combination of electrode, substrate and counter electrode. R_g and C_g resistance and capacitance of the air gap; the Zener diode Z takes account of the breakdown voltage of the gas discharge. R_d and C_d resistance and capacitance of the dielectric layer. R_p and C_p resistance and capacitance of the paper. R_c and C_c are the effective resistance and capacitance of the contact impedance. All resistance and capacitance values are for unit area.

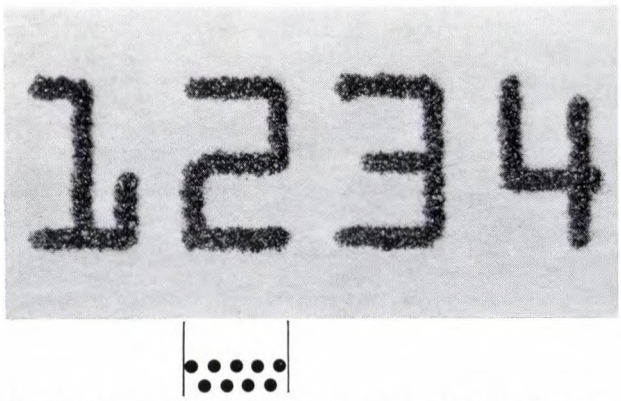


Fig. 7. OCR-A characters, printed with 7 electrodes per mm. This series of numerals has been designed for simple recognition in a reading machine. The electrode configuration is shown beneath the figure 2; four characters can be printed per cm of line length.

Such a low paper resistance can only be achieved by taking special measures to increase the conductivity. The best electrical properties are obtained by introducing finely divided carbon into the paper, but this does have the disadvantage of darkening it. If it is essential to have white paper, the only way of increasing the conductivity is by ionic conduction, i.e. by applying ionized salts. The conductivity obtained in this way depends on the degree of dissociation of the salts, and hence on the humidity of the paper and consequently on the ambient humidity.

As well as the resistances and capacitances discussed above, the equivalent circuit also includes the transition to the counter electrode, which is represented by R_c and C_c. For charging with a single electrode these can be taken together with R_p and C_p, the corresponding quantities for the paper.

Charging with a row of pin electrodes

As already indicated in the introduction, the electrostatic printing techniques have a very wide field of application. In one particularly adaptable kind of system, an image is produced by printing separate points with a fixed electrode system. The preferred arrangement for this is a row of closely spaced pin electrodes. These can be spaced as closely as ten per millimetre.

For printing alphanumeric characters, which are generally printed with a point matrix, 4-6 pins per mm are usually sufficient; for the OCR-A characters (*fig. 7*), however, seven pins per mm are necessary. The shapes of the various characters are stored in a coding memory, e.g. in a circuit used as a 'read only' memory (ROM) as patterns of dots arranged in horizontal rows. A line of characters is printed as a number of rows of dots formed by calling up in succession the rows of dots for all the characters in the line. At the beginning of each row of dots the correct rows for all the characters have to be chosen from the coding memory.

With a page of A4 format 800-1000 points are required for the length of a line. To enable all these points to be separately energized with a voltage of 400-500 V, the same number of switches is required. However, since a gas discharge first breaks down above a particular voltage, it is possible to make use of a coincidence circuit. The number of switches required is then considerably less.

An initial version of a coincidence circuit is shown in *fig. 8*. Each pin electrode is connected via two

[8] J. Schramm and K. Witter, Appl. Phys. **1**, 331, 1973.

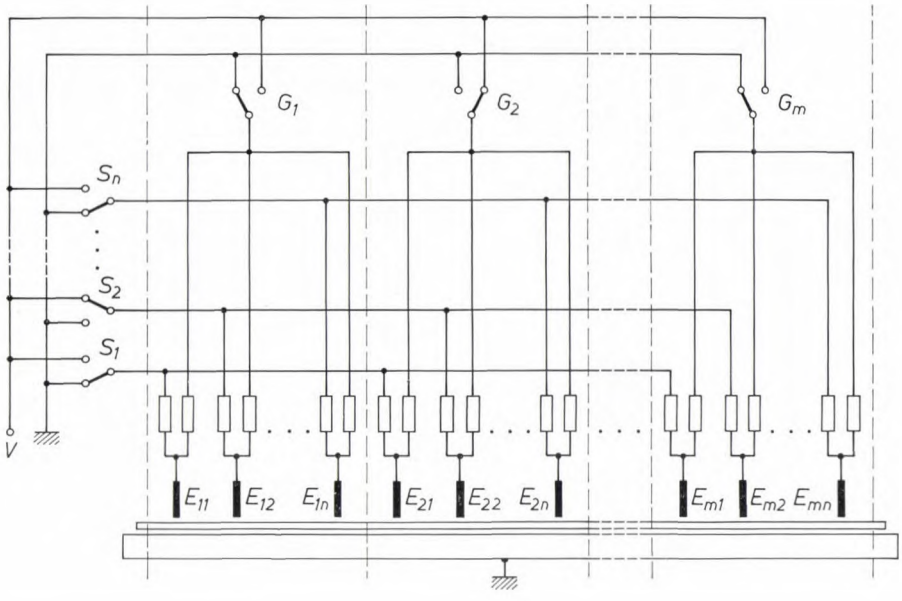


Fig. 8. Diagram of a coincidence circuit for energizing $n \times m$ electrodes. The electrodes are divided into m groups of n electrodes. Each electrode is connected via two resistors to one of the group switches G and to one of the character switches S. Only those electrodes that are connected to the voltage line through both switches will be at the voltage V; in other cases the voltage is $\frac{1}{2}V$ at most. The voltage V is chosen such that V is greater and $\frac{1}{2}V$ smaller than the breakdown voltage of the discharge between an electrode and the substrate.

resistors to two switches, the character switch and the group switch. These switch a voltage V, which is a little lower than twice the breakdown voltage for the discharge: V would thus have a typical value of 700 V. Only those electrodes that are connected to the voltage line through both switches will in fact be at this voltage. If only one of the two switches is connected to the voltage line and the other is connected to earth, then because of the potentiometer effect only half the supply voltage will appear at the electrode. The gas therefore does not break down, and the electrode applies no charge to the paper.

In another method, now widely used by several manufacturers, a group of electrodes is selected with the aid of a counter electrode consisting of separate sections, each of which can be independently switched [9]. The spacing between the electrodes is so small (about 0.1 mm) that it is not possible to separately use two electrodes on either side of the boundary between two sections of the counter electrode. This is because the potential does not change sharply at the boundary, but changes gradually over a distance of a few millimetres (*fig. 9*). The width of this zone depends on the resistance of the paper and the contact impedance between paper and counter electrode. If not too high a voltage is selected for charging, the images obtained are blurred at the edge of the counter electrode in use at that moment. At higher voltages the edge quality is good, but characters of the previous or following group may appear as shadows in the text.

The situation in the boundary region between two counter electrodes is shown in *fig. 10*. We shall assume that one electrode is at a voltage $+V_0$, the other at $-V_0$; the distance between the two electrodes is $2b$. The contact conductivity per unit area is represented by G_e (dimensions $\Omega^{-1}\,\mathrm{cm}^{-2}$) and the resistivity of the paper by ϱ_p (dimensions Ω cm). If we now assume that $1/G_e \gg \varrho_p d$, where d is the paper thickness — a condition always fulfilled in practice — a very simple relation can be derived for the voltage variation $V(x)$ in the substrate [10]:

$$\frac{\mathrm{d}^2V}{\mathrm{d}x^2} = \frac{\varrho_p G_e}{d}\left\{V(x) - V_0\right\}.$$

Putting $\varrho_p G_e/d = \mu^2$, we have for $x \geqslant b$ the solution

$$V(x) = V_0\left\{1 - \frac{\exp(-\mu x)}{(1+\mu b)\exp(-\mu b)}\right\},$$

the graph of which has already been shown in fig. 9. The points $x = \pm b$ are important here; at these points the potential $V(\pm b)$ in the substrate must be as close as possible to the electrode potentials $+V_0$ and $-V_0$.

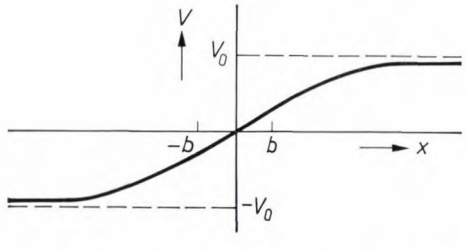


Fig. 9. The potential V on the substrate as a function of the distance x, at the boundary between two counter electrodes. One electrode is at a voltage $+V_0$, the other at $-V_0$; the distance between the two electrodes is $2b$.

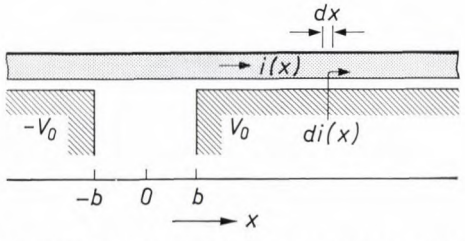


Fig. 10. Diagram relating to calculation of the potential variation along the substrate in the boundary region between two counter electrodes. The current in the substrate at the location x is $i(x)$; the increase in the current over the distance $\mathrm{d}x$ is $\mathrm{d}i(x)$.

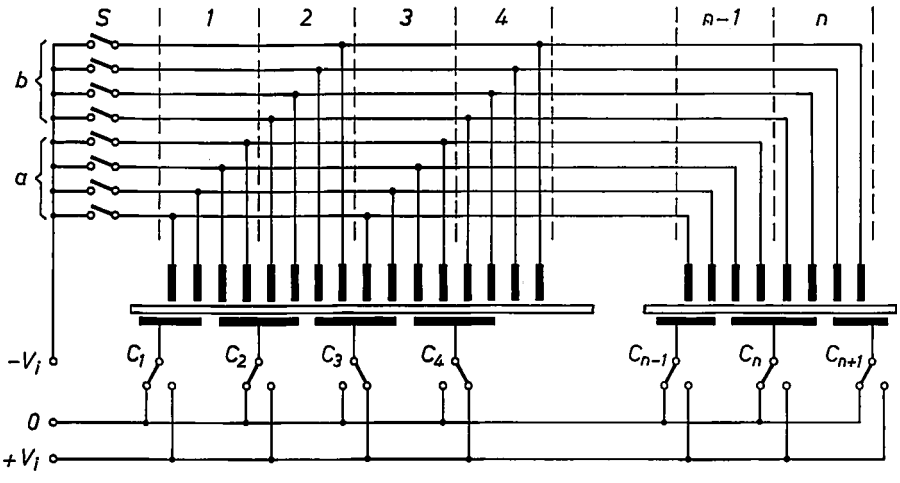


Fig. 11. Diagram of the circuit with separately energized counter electrodes. The groups *a* and *b* of the character switches *S* are used alternately. If for example one of the electrodes in group *3* has to be used, then the counter-electrode switches C_3 and C_4 must also be connected. This ensures that the potential of the substrate opposite group *3* is everywhere high enough for breakdown of the discharge. Opposite simultaneously connected electrodes in the groups *1* and *5*, however, the potential is not high enough for breakdown.

Since

$$V(b) = V_0 \frac{\mu b}{1 + \mu b},$$

we must therefore have $\mu b \gg 1$. In practice μ will be about 1 mm^{-1}, so that special precautions will have to be taken for counter electrodes for which $b < 1$ mm.

These difficulties can be avoided by arranging that the boundary between two groups of pin electrodes is opposite the centre of a counter electrode, and then always energizing two adjacent counter electrodes simultaneously (*fig. 11*). Although this does require separate control for adjacent electrode groups, the edge effects mentioned above are eliminated if the groups are sufficiently wide. In comparison with the circuit with the resistance network the switches now only have to handle half the supply voltage; on the other hand in the most favourable case $\sqrt{2}$ times as many switches are required.

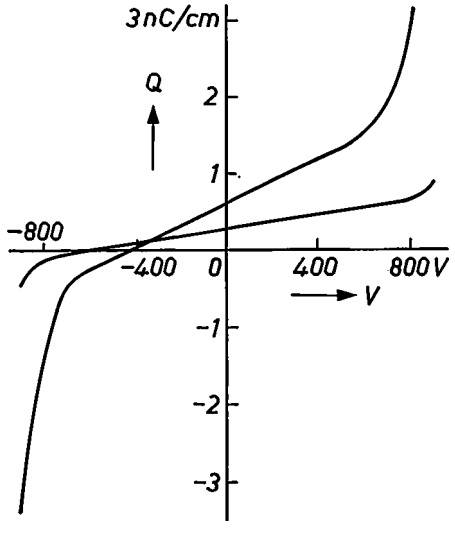


Fig. 12. The charge Q per unit of line length as a function of the applied voltage V for charging by direct contact between electrode and substrate. The two curves apply for polycarbonate sheet of different thicknesses, and hence different capacitances per unit area.

Charging by direct contact between electrodes and substrate

The previous section described charging by a self-sustained gas discharge between the electrode and the substrate. Such a discharge must always extend over a small region in order to exist, and therefore does not give a true image of the electrode on the substrate. The charge spot on the substrate is appreciably larger than the electrode, particularly for small electrodes. The smallest charge spot that can be produced in this way has a diameter of about 0.1 mm. To achieve a high resolution with an electrostatic printing process another method of charging must therefore be used, without the self-sustained discharge. This also implies that the electrode voltage must be lower than the 380 V corresponding to the minimum of the Paschen curve.

We have worked with a direct contact between the electrode and substrate [11]. This can only be done when both electrode and substrate are very smooth, so that the pressure of the electrode against the substrate can be fairly high. Only certain very smooth sheet plastics can then be used, and the electrodes must be made of extremely hard material that will take a polish, but must also be a sufficiently good conductor. The substrate must be fairly thin to give a sufficiently high capacitance per unit area, so that sufficient charge can be applied to the substrate even with low voltages. In our experiments we used 'Mylar' sheet plastic, with one side metallized to produce a counter electrode.

Fig. 12 effectively summarizes the results of experiments in which pin electrodes were pressed against sheet plastic with a pressure of about 5×10^6 N/m^2. In the central region there is a linear relation between the

[9] U. Rothgordt, West German Patent No. 1 946 815.
[10] U. Rothgordt, Philips Res. Repts. 29, 139, 1974.
[11] U. Rothgordt, K. Witter and H. D. Hinz, to appear shortly in IEEE Trans.

applied voltage V and the charge it produces per unit length of the written trace. Dividing by the width of the trace (100 μm) gives the surface charge density σ. At $V = 0$ friction alone gives a certain charge density σ_0. In practice it is found that σ_0 is constant for a given combination of electrode and substrate material; it is therefore possible to compensate for σ_0.

The central portion of the curves can be simply described by $\sigma = \sigma_0 + (\Delta\sigma/\Delta V)V$. It is found that $\Delta\sigma/\Delta V$ is exactly equal to the capacitance per unit area of the substrate. This means that here, unlike the case of charging with a gas discharge, the full applied voltage contributes to the supply of charge. Another important feature is that charge is only applied where electrode and substrate actually touch. Fig. 12 shows that at higher voltages strongly nonlinear behaviour appears. This is due to a gas discharge between the sides of the electrode and the substrate; it is associated with a distinct broadening of the trace. For good resolution these higher voltages should therefore be avoided.

In recording a charge image each point of the complete surface of the substrate must come into contact with an electrode in turn. Because a fairly high contact pressure is required electrode arrays cannot be used here, and only a limited number of electrodes can be allowed to make simultaneous contact with the substrate. The device whose principle is shown in *fig. 13* satisfies this requirement. The most important component is a rapidly rotating cylinder. The substrate covers a part of the circumference of the cylinder. An insulated electrode is mounted in the cylinder wall, and protrudes 100 μm beyond the cylinder surface. The cylinder is given a small axial displacement after each revolution, so that the electrode describes a raster on the stationary substrate. By varying the electrode voltage appropriately a charge image can be written on the substrate. The situation in the immediate neighbourhood of the electrode is shown in *fig. 14*. Charge spots about 10 μm in diameter can be produced with this device. An experimental microprinter based on this principle has been produced, which can print characters reduced by a factor of 24. Such printers are generally referred to as COM (for Computer Output on Microfilm) printers. Compared with other COM printers the one described here has the advantages of being smaller and less expensive; it can print at the same rate as fast mechanical printers.

A virtual feature of such equipment is the behaviour of the 'stylus' that makes contact with the dielectric sheet. Wear is unavoidable here. The hardest materials, such as diamond or sapphire, cannot be used here, because of their high electrical resistivity. It has however been found that whiskers of SiC (prismatic, single-crystal filaments of diameter about 20 μm) give sufficient conductivity for this application. Because of the single-crystal structure the wear of the whiskers is small, and also very uniform. Since no change in the shape or size of the whisker surface occurs in use, up to 5×10^{10} points can be printed with a single electrode before any deterioration in quality is observed. This figure corresponds to about 10^5 printed pages.

Electrostatic grey-scale printing

For grey-scale printing it is necessary to vary the charge applied to the separate points. This is difficult with the methods that use a self-sustained discharge.

We therefore directed our work towards a variable charge source. An important condition here is that the charging process must take place under normal atmospheric conditions, so that the substrate does not have to be inside an evacuated enclosure, as would be the case in charging with an electron beam.

These requirements can be met with various types of gas discharge, such as spark, arc, glow and corona discharges. For simplicity, the obvious one to choose

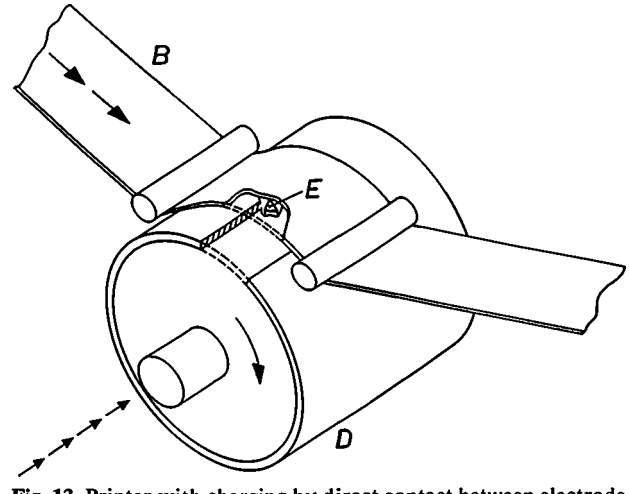


Fig. 13. Printer with charging by direct contact between electrode and substrate. The 'Mylar' belt *B*, on which the image is printed, lies against the surface of the rapidly rotating drum *D*. The electrode *E* protrudes about 100 μm beyond the surface of the drum. The drum is slowly displaced, continuously or in steps, along its axis (one step per revolution). In this way charge is applied to the substrate line by line. When an image is complete the belt has to be moved across an image width.

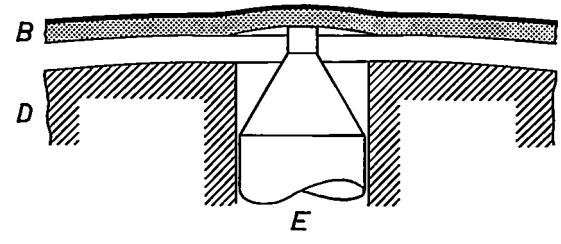


Fig. 14. Cross-section of the electrode of the printer of fig. 13. The rapid rotation of the drum *D* increases the air pressure between the surface of the drum and the belt *B*. This lifts the belt from the drum slightly. The little 'bulge' in the substrate where the electrode *E* makes contact ensures sufficient contact pressure.

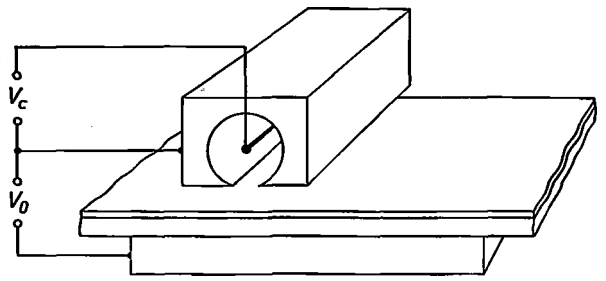


Fig. 15. Cylindrical corona discharge as source of charge carriers. A corona discharge (corona voltage V_c) is maintained between a cylindrical electrode and an axial wire. The charge carriers are drawn towards the substrate by a voltage V_0 between the outer electrode and a 'drain'.

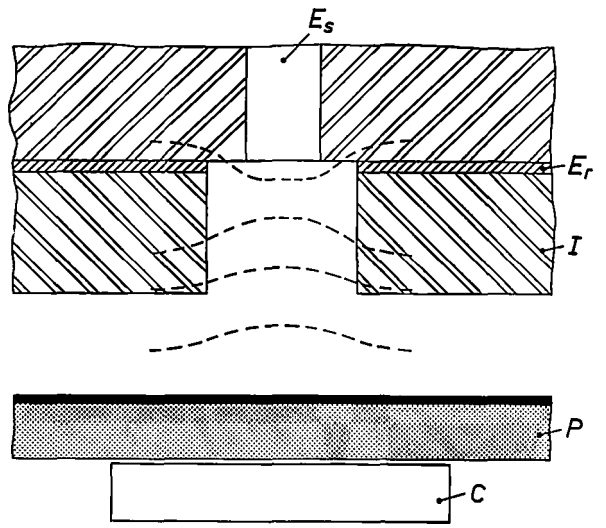


Fig. 16. Electrode configuration for grey-scale printing with a spark discharge. Ions produced in a spark discharge between the pin electrode E_s and the ring E_r are accelerated in the direction of the substrate P by the field from the drain C. The dashed lines indicate the field configuration between the various electrodes, which gives a contraction of the ion beam. The charge spot produced is smaller than the diameter of the cylindrical hole in the insulator I. The degree of contraction depends on the voltage on the electrode C.

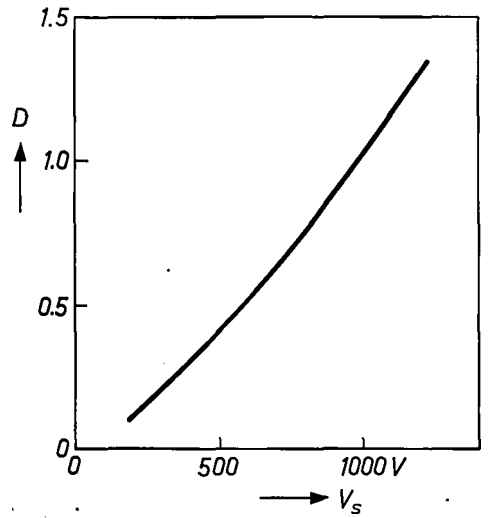


Fig. 17. The optical density D as a function of the drain voltage V_s for print produced with the arrangement of fig. 16.

would be the corona discharge. However, with a simple cylindrical corona discharge as in the arrangement of *fig. 15* the current density is too low to give a practical printing speed. It has been reported [12] that the current density can be increased by using a pin electrode in combination with a special gas in the discharge region.

In our laboratories we have used a spark discharge [13]. The charge carriers from the discharge are focused to a narrow beam by a diaphragm functioning as an electron lens. A cross-section of the discharge region is shown in *fig. 16*. This construction can be obtained by placing in front of the pin electrode a thin sheet of ceramic, metallized on one side, with a hole in it slightly greater than the pin and concentric with it (hole 250 μm, pin 150 μm). The perforated metallization serves as a ring electrode and is separated from the pin by a gap of 50 μm. The substrate is located up to a millimetre away from the spark electrodes. The counter electrode, behind the substrate, is here more appropriately called a 'drain'.

The similarity to the electrode of fig. 4 is however only in appearance. The ring electrode is connected to earth, while high-voltage pulses of 1 to 1.5 kV and pulse length several μs are applied to the pin electrode. This produces a brief spark discharge between the electrodes. The discharge produces large numbers of charge carriers, and these are drawn off by applying to the drain a voltage pulse of opposite sign to that on the pin and of somewhat longer duration.

The field distribution obtained ensures that for constant pulse height and duration a constant quantity of charge carriers arrives at the substrate in a spot roughly corresponding to the hole in the insulator. The diameter of the spot and the charge density depend on the diameter of the hole and the height of the two pulses.

By varying the height of the voltage pulse on the drain the charge density can be modulated, thus giving different shades of grey when the charge image is developed [14]. The relation between the optical density and the voltage of the drain is shown in *fig. 17*. From the information presented it follows that grey shades can be satisfactorily printed with the electrode construction described. With this arrangement we have succeeded in obtaining 15 shades of grey distinguishable to the eye at a resolution of 10 points per mm.

The spark-electrode construction is of interest not only because of the charge control it offers, but also because of the much larger distance permissible between electrode and substrate without adversely affect-

[12] W. Simm, 4. Int. Kongreß für Reprographie und Information, Hannover 1975, p. 26.
[13] G. Krekow and J. Schramm, Feinwerktechnik + Micronic 77, 219, 1973.
[14] G. Krekow and J. Schramm, IEEE Trans. ED-21, 189, 1974.

ing the magnitude or sharpness of the dots of charge. In writing a charge image on the rotating drum of an image-transfer machine, for example, it is difficult to keep all the pins of a row of electrodes at an accurate spacing of 10 μm from the surface of the drum. Charging with spark electrodes, which will operate at a spacing of a few tenths of a millimetre from the drum, offers a solution.

Developing and fixing a charge image

The developing process, whose name is borrowed from photography, is not a chemical reaction here. It consists in bringing the substrate on which the charge image is written into contact with a fine dye powder, usually called the toner. The toner particles are held in position at the charged places on the substrate by electrostatic forces. The development only proceeds successfully when the toner particles are brought into contact with the substrate in a relatively low concentration; with higher concentrations the toner would stick to the substrate everywhere because of molecular forces. The toner particles are therefore either suspended in a fluid, or mixed with a much coarser powder, the carrier.

Both methods of developing have their own advantages and disadvantages. Developing with a suspension is chiefly recommended for high-resolution printing processes, like the method described earlier in which there was direct contact between electrode and substrate. If a powder developer is used, it is much easier to transfer the developed image to another substrate: we shall return to this transfer process later. For developing the charge images obtained with a gas discharge we have given preference to powder developers.

The developing process is shown schematically in *fig. 18*. The powder developer consists of about 2% toner and 98% iron powder, which is the carrier. Friction between the two kinds of particles gives them opposite electrical charges.

Except at the edges of the charged areas, the field from the negative charge on the dielectric layer is almost entirely *inside* the layer, and does not emerge from it (fig. 3). If the conducting iron powder is brought into its surroundings, however, the whole field pattern changes and a field does appear outside the layer. If the field-strength is high enough, toner particles will be pulled away from the iron powder and will be attracted to the charged areas of the substrate surface. A sufficient degree of development is obtained at a charge density of 2×10^{-4} C/m², while 5×10^{-4} C/m² is required for complete saturation.

The transfer of the toner particles from the iron powder to the substrate surface is a charge transport and is therefore associated with a current. To maintain the original potential differences the circuit for this current must be completed, as shown schematically in fig. 18. In general there will also be a limited amount of toner deposition on the uncharged parts of the substrate: this can be prevented by including a voltage generator (voltage up to about 20 V) in the short-circuiting connection.

Fresh toner will obviously have to be added to the developer mixture from time to time to make up for the toner that has been used. It is however more important to ensure good homogeneity of the developer mixture. A number of methods have been investigated for this purpose; the most important ones are the cascade process and the magnetic-brush method. In the cascade process the substrate carrying the charge image is tilted and sprinkled with the developer mixture. The mixture flows downwards over the substrate, leaving a thin layer of toner at the charged places. The magnetic-brush principle can be applied in many ways. In one version, a round bar magnet is used, mounted inside a closed non-magnetic tube. When magnet and tube are dipped in a mixture of toner and iron powder, a 'brush' is formed at the end of the tube. When the brush is moved over the substrate toner remains behind at the charged places. The developer arrangement pictured in fig. 2, is a variant of this principle. Magnetic-brush development gives better image cover, since more toner can be applied to the charged places than with the cascade process.

A particularly suitable and easily constructed solution to this problem is our eddy chamber, shown in *fig. 19*. The iron powder used here as carrier for the

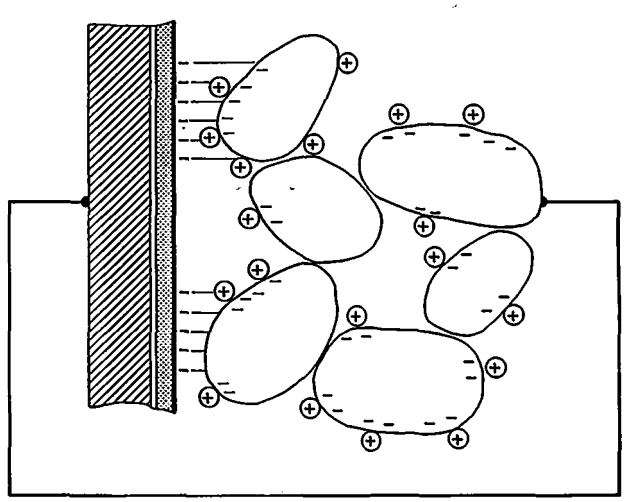


Fig. 18. Powder developing process with a mixture of coarse iron particles and fine toner particles. The two kinds of particles become oppositely charged by friction between them. A conducting path between the iron particles and the back of the substrate causes the lines of force from the charge to emerge from the substrate, as was shown in fig. 3.

toner is drawn against the substrate by a magnet behind the paper. If the paper moves upwards, powder mixture is carried along by friction with the paper. Because of the combined effects of the magnetic forces and gravity very strong eddies are produced in the powder, and these ensure effective homogeneity. A developer of this type gives satisfactory results with paper speeds up to 5 m/s. The principle is less useful for paper speeds below 10 cm/s, since the eddy effect is not then strong enough. In such a case one form or another of magnetic brush would be preferable, e.g. the arrangement shown in fig. 2; this has helical ridged rollers to homogenize the powder mixture. The roller speed must be matched to the speed of the paper.

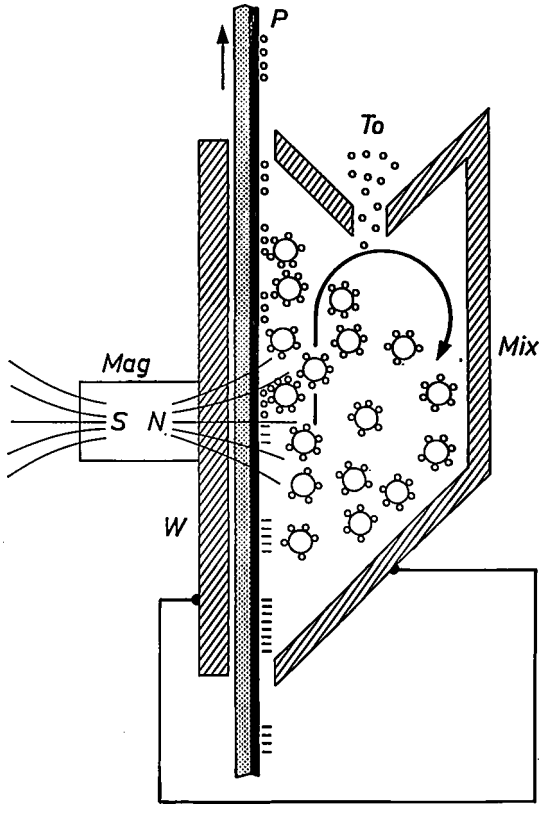


Fig. 19. Eddy chamber for developing electrostatic printing. The paper strip *P*, coated with a dielectric layer that carries the charge image, moves vertically upwards. The back of the paper runs against the chamber wall *W*. On the other side is the mixing chamber *Mix*, containing a mixture of toner and iron powder. This mixture is drawn against the paper by the magnet *Mag*. The powder is carried upwards with the paper because of friction; this initiates eddies in the powder, which ensures thorough mixing of toner (*To*) and iron powder.

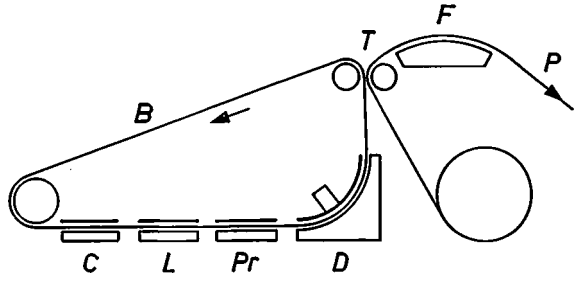


Fig. 20. Diagram of an electrostatic printer with a transfer surface to transfer the toner image to ordinary paper. The transfer surface, the endless belt *B*, runs anticlockwise; it is printed at *Pr* and developed at *D*. At *T* the toner image is transferred to the paper *P* (the final surface) and is fixed at *F*. The transfer surface is cleaned at *C* and the charge image is neutralized at *L*. A new charge image can then be applied to the transfer surface.

The final stage in the direct printing process is fixing the toner, now relatively weakly attached to the substrate. In developing with a toner suspension adhesives are added to the fluid; these provide the bond between toner and substrate when the fluid has evaporated. In dry-development processes are used in which the particles are bonded to the substrate by the effect of mechanical pressure or by heating.

The transfer of electrostatically printed images

So far we have only examined the direct printing process, in which the charge pattern is applied to the actual surface to be printed. This surface was therefore provided with a dielectric coating. There are also printing processes in which the charge image and the developed powder image are produced on a transfer surface, and the powder image is then transferred to ordinary paper [15]. The transfer surface can then be used to give many prints of the image. An endless belt of dielectric material is often used as the transfer surface, or a drum with a dielectric coating. This process is of particular interest for the high-speed print-out of computer results, an application in which the cost of the paper is important and more than one print of the same text is usually required. With these requirements in mind we have studied more closely the transfer of a powder image.

The transfer, which could take place in the arrangement shown in the diagram of *fig. 20*, can be briefly described as follows. The transfer surface with the powder image is brought into contact with a final surface (paper) that runs at the same speed as the transfer surface. A potential difference of the appropriate polarity is now applied between the two surfaces; this removes the charged toner particles from the transfer surface and transports them to the final surface. The potential difference required may be a few hundred volts; it must not however be greater than the breakdown voltage of air. There are no particular requirements for the electrical properties of the final surface. To ensure that the toner particles do indeed remain attached to the final surface, it is important that the two surfaces are separated in the electric transfer field.

[15] U. Rothgordt, West German Patent Application No. 1 272 316.
H.-D. Purps and U. Rothgordt, Int. Conf. on Communications, Montreal 1971, p. 15-14.

The charge image on the transfer surface is preserved when the powder is transferred, so that development and transfer can be repeated if desired, to give one or more copies.

Before new information is introduced all the remaining toner must be removed from the transfer surface and the charge image must be neutralized. Residual toner can be removed with a cylindrical brush, a 'vacuum cleaner' or a squeegee. The neutralization of the charge image is in fact a reversal of the charging process and is also carried out with a gas discharge. In this process the transfer surface runs close to a number of sharp metal knife-edges, to which an erase voltage is applied. The magnitude of this voltage is such that at places where there is no charge on the surface, no discharge is produced. At the charged areas, however, the erase voltage and the potential due to the surface charge combine to produce breakdown, which erases the charge image.

As mentioned above, either drums or endless flexible belts are suitable as the transfer surface. They must have the same electrical properties as formulated earlier for special paper, and must be sufficiently stable mechanically to permit 10^4 to 10^6 print cycles to be run without significant loss of quality.



Fig. 21. Text from an electrostatic printer. The maximum printing speed of the printer used here is 20 000 lines/min.

Applications of the electrostatic printing process

The electrostatic printing process can be very adaptable and is therefore of interest in cases where various very different kinds of data have to be recorded rapidly. The only limitation is that it does not give print instantaneously (like a typewriter). The main advantages derive from the speed of the actual printing process, which is high, even compared with other non-mechanical processes, and in the silent operation of the process.

The difference between the examples now to be described is not so much in the actual printing process as in the pretreatment used to convert the information to be recorded to a form suitable for the electrostatic printing process.

For printing alphanumeric characters a fixed comb of pin electrodes is generally used. The method of printing letter characters with a matrix, typically of seven by nine points, has been given earlier. Some text printed in this way on a machine that can reach speeds of 20 000 lines/min (paper speed 1.4 m/s) is reproduced in *fig. 21*.

Another interesting application is the combination of an electrostatic printer with a computer visual display. In these displays a television-type raster is often used for reproducing the information, with alphanumeric characters imaged by a dot matrix. This means that the equipment associated with such a

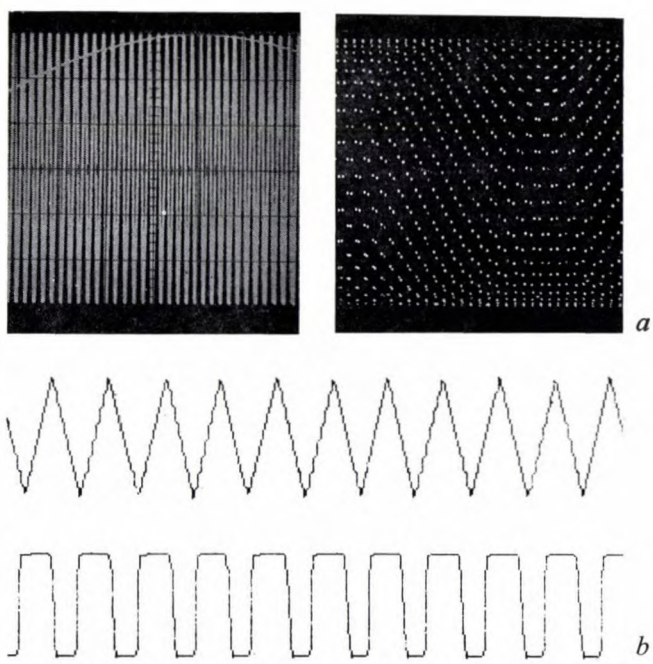


Fig. 22. *a) Left:* A.C. waveform, shown on an oscilloscope screen. *Right:* The same waveform, but now reproduced by sampling at regular intervals in time. The coherence between successive points in time can now no longer be seen (unless the figure is viewed in such a way that the height is foreshortened).
b) Electrostatic prints of sampled values, with the coherence between successive points restored by printing vertical lines between two sampling values.

display can easily be modified to produce an electrostatically printed text. In this way the total content of the display can be put on paper in less than five seconds.

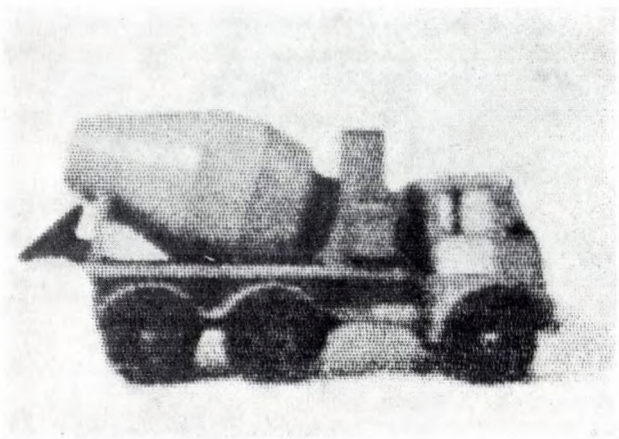


Fig. 23. Enlargement of a grey-scale image obtained by electrostatic printing. The original image was 21 × 35 mm.

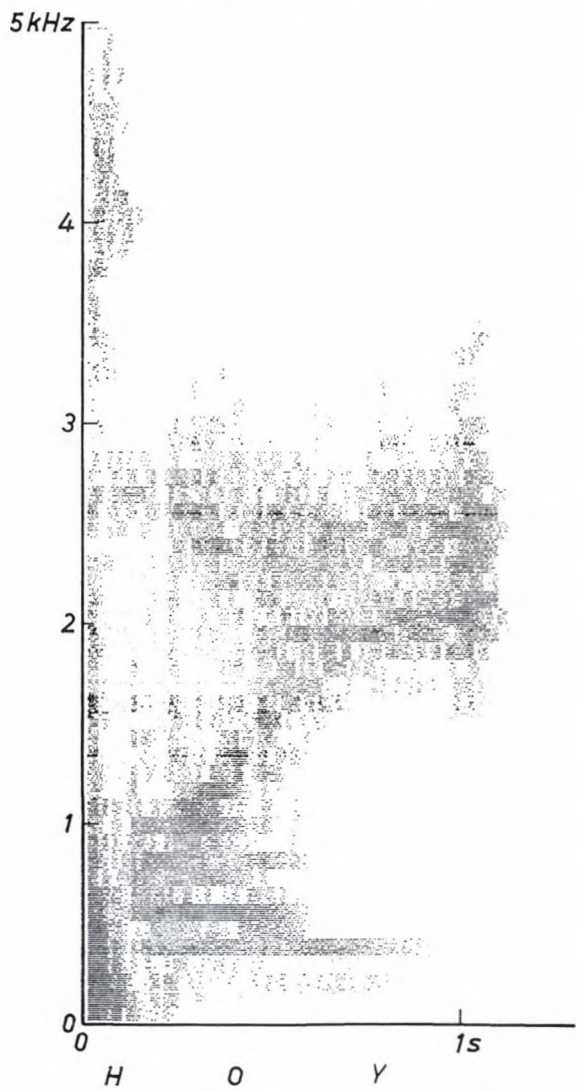


Fig. 24. Time-dependent acoustic spectrum ('visible speech diagram') for the spoken sound 'Hoy'.

An example of an electrostatic printer that will 'draw' is a recorder. The analog signal to be recorded must be sampled and digitized at regular intervals. Each of the values obtained must then be applied to a particular pin in an electrode comb. The points printed sequentially in this way must approximate as closely as possible to an uninterrupted line. If, however, optimum use is made of the analog-to-digital converter and the available time per sampling, it cannot always be guaranteed that an uninterrupted line will be produced. This can lead to moiré effects, as *fig. 22a* shows. The original coherence of the points is then no longer obvious, because points that follow one another in time and are almost on the same vertical line will be much further apart than points belonging to different periods of the signal, but at similar horizontal levels. To restore the coherence between successive points without raising the sampling rate of the analog-to-digital converter, we print not only the point for the signal value at a particular instant, but also, at the same time, all the points between this value and the last recorded point. The result of this processing is shown in fig. 22*b*. The desired coherence is restored.

Another field in which electrostatic printers are of increasing importance is that of facsimile systems. In these systems a document at the transmitting station is scanned point by point, the density of each point coded, sent over a telephone line, and a point of the desired density printed at the receiving station. In the simplest and most common arrangement only two levels are distinguished: black and white. An electrostatic printer with a comb of pin electrodes is eminently suitable for such a system; a particular electrode can be selected rapidly and virtually at will. This is of special interest because modern facsimile systems use an optimum run-length coding system, in which the information is not transmitted point by point, so that a printer that follows step by step is less desirable. The speed of an electrostatic printer can readily be adapted to handle the massive flow of coded data because of the short time necessary per point (about 10 μs). This is not however the place to go into further detail about these coding systems [16].

Besides the transmission of black-and-white images, such as text or line drawings, it may also be necessary to transmit shades of grey, as in the transmission of satellite photographs. The grey-scale printer described earlier is very suitable for this purpose. *Fig. 23* shows an image produced with such a printer [14].

As the final example, the use of an electrostatic printer as a rapid *x-t* plotter will be mentioned. In this case

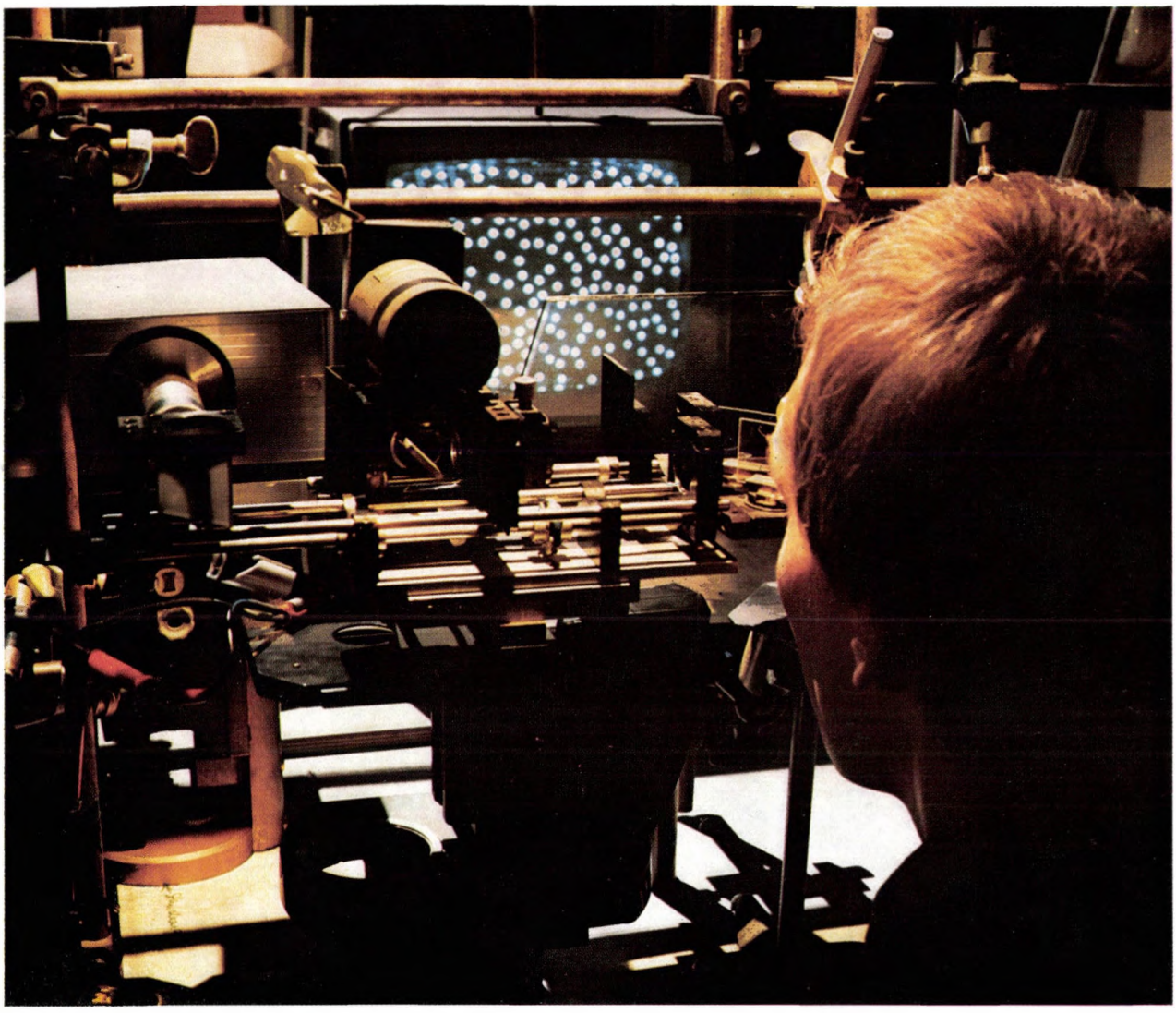
[16] But see for example H. Meyr, H. G. Rosdolsky and T. S. Huang, IEEE Trans. **COM-22**, 826, 1974.
U. Rothgordt, Electronics Letters **11**, 101, 1975.

the printer, which has already been shown in fig. 1, is connected to a computer, and the system thus formed provides real-time spectra of acoustic signals [17]. The technique of 'Fast Fourier Transformation' for this application has been known for some time; it can be successfully used in obtaining spectra at scanning frequencies up to 10 kHz [18]. The Fast Fourier Transform gives the acoustic energy present in a large number of frequency intervals of width Δf during a fixed time interval Δt. The width of the frequency intervals depends directly on the rate at which the complete spectrum is scanned. The time-dependent spectrum produced in this way consists of the separate regions $\Delta t \times \Delta f$, with a signal amplitude indicated for each region. With an electrostatic printer this spectrum can be recorded by showing each region as a matrix of 64 locations, with more or fewer locations occupied by black points, depending on the amplitude to be recorded. With 4 points per mm each matrix occupies a 2-mm square. *Fig. 24* gives an example of such a spectrum.

Summary. Lines and alphanumeric characters can be drawn by applying a charge distribution corresponding to the desired pattern to a dielectric sheet or to coated paper. The charge can be applied by producing a gas discharge between a pin electrode above the substrate and a flat electrode below it; it is also possible to move a thin conducting pin over the substrate and to apply a voltage between the pin and the back of the substrate. Another possibility is to produce a spark discharge between two electrodes above the substrate and to draw charge carriers towards the substrate by a voltage on an electrode placed below it. The smallest charge dots that have as yet been produced with a gas discharge have a diameter of about 100 μm; with direct contact diameters down to 10 μm have been reached. The charge image is then developed by bringing it into contact with a mixture of dye powder (toner, particle size 5-30 μm) and iron powder (100-400 μm). The iron powder flows freely because of its fairly large particles, ensuring an even distribution of toner particles over the charge image. The conducting iron particles also have the function of extending the electrostatic field due to the charge image to the outside of the dielectric layer. It is also possible to use a suspension of toner in a suitable fluid for development. Finally, the image must be fixed on the substrate. The main applications of the electrostatic printing process are in data processing, where the high speeds attainable (up to 20 000 lines per minute) are of great interest.

[17] U. Dibbern, Int. Conf. on Speech communication and processing, Boston 1972, p. 412.
U. Rothgordt, 8th Int. Congress.on Acoustics, London 1974, p. 721.
[18] A. V. Oppenheim, IEEE Spectrum 7, No. 8, p. 57, August 1970.

The measurement of visual conspicuity



Man only perceives a small part of what is to be seen around him. The choice of viewing direction is an initial, rough selection. The observer's attention then determines the part that is consciously seen of what falls within the field of view. What does this being seen depend on? We all have some idea about this: something is noticed because it differs from its environment in qualities such as shape, colour or size. An understanding of selective perception in vision is desirable for such purposes as the designing of traffic-signal systems and the appraisal of inspection and checking tasks.

Experiments aimed at providing a fundamental insight into the process of visual perception have to be based on a highly simplified observation situation. For this purpose we have chosen a test pattern consisting of a large number of randomly situated light spots of equal size and brightness against a dark background, to which one spot of deviating size is added. The question now is what determines whether the deviating spot is noticed at a single glance. It is found that around the fixation point a 'conspicuity area' can be indicated within which the deviating spot must lie if it is to be noticed. The subject is instructed to look fixedly at a certain point on a monitoring screen, on which the test pattern is briefly displayed (for a time of 75 ms), and then to indicate where the deviating spot was situated. The time of 75 ms was chosen to eliminate the influence of eye movements, which do not take place uniformly but in intervals of 150-500 ms. Because of the influence of the structured background the conspicuity area proves to be much smaller than the total visual field (*fig. 1*). The size of the conspicuity area is determined

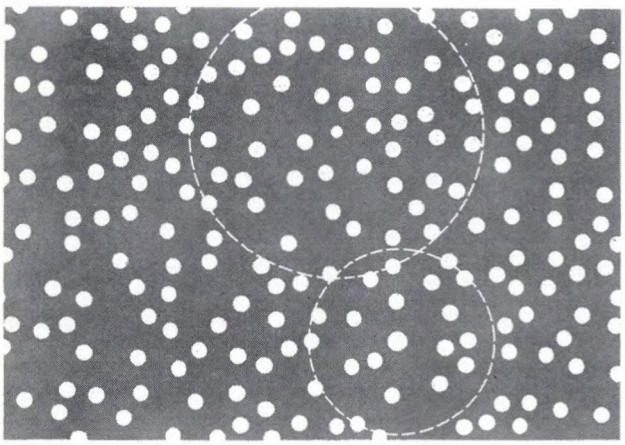


Fig. 1. One of the test patterns used in the investigation described here. In this case there are two spots whose diameter differs from that of the background spots. The dashed circles approximate to the conspicuity areas of the deviating spots, for normal reading distance (25 cm).

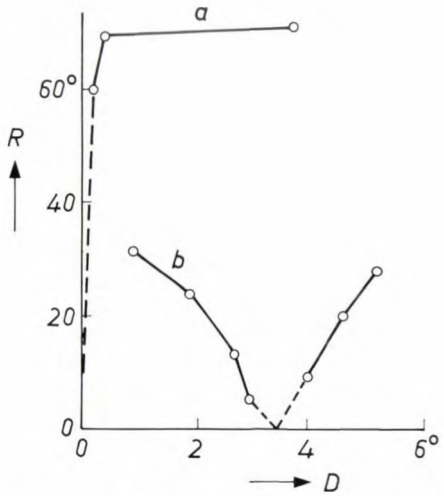


Fig. 2. The 'radius' R of the conspicuity area of a spot as a function of the diameter D of the spot. Angular dimensions are given. Curve a relates to a single bright spot against a uniformly dark background, curve b to a bright spot against a background of randomly distributed spots of equal brightness with a diameter of 3.5°.

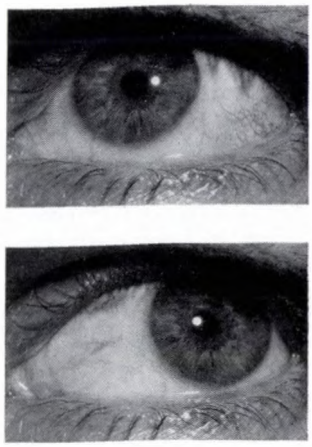


Fig. 3. Displacement of the mirror image of a fixed light point given by the cornea of the eye for two extreme directions of the visual axis. The position of this mirror image is found to give a useful measure of the direction of the visual axis. An infrared light source was used in these experiments, to prevent the subject from being affected by glare.

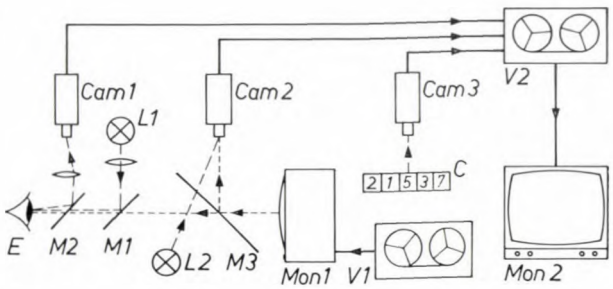


Fig. 4. Diagram of the arrangement for studying visual search. With the aid of a video recorder *V1* a test pattern is displayed on a monitor *Mon1*, which is viewed by the subject (*E* eye of subject). To record the movements of the eye a narrow beam of infrared radiation from a lamp *L1* is directed through a semi-reflecting mirror *M1* on to the cornea of the eye. The beam reflected from the cornea passes through the semi-reflecting mirror *M2* to the video camera *Cam1*. The test pattern is picked up, via *M3*, by camera *Cam2*, as also is the light from lamp *L2*, with which the subject indicates that he has found the deviating spot. Time measurements are made by a digital clock *C*, which is picked up by the camera *Cam3*. The signals from the three video cameras can be collectively recorded on video recorder *V2* and observed — while the experiment is taking place or later — by the experimenter on monitor *Mon2*. The magnification of *Cam1* is chosen in such a way that the position of the cornea reflection on *Mon2* corresponds to the subject's point of fixation on *Mon1*.

by the extent to which the spot searched for deviates from the background. Factors of importance here are the background structure and the difference in size and brightness between the deviating spot and the background spots (*fig. 2*). The concept of the conspicuity area allows a good description of the visual search behaviour of an observer to be obtained. It is necessary, however, to distinguish the conspicuity area from the larger area within which the detail will be discerned by an observer who knows its location.

The title photograph shows the experimental arrangement used in our investigation of visual search behaviour. The subject looks at a test pattern consisting of an arbitrary arrangement of spots in which one spot differs from the rest. The observation time is not fixed, but the observer searches until he finds the deviating spot. The position of the subject's head is fixed, to enable the eye movements to be recorded (*fig. 3*). The subject gives a signal as soon as he finds the deviation, so that the time he needed to find it is known. All this information is recorded on magnetic tape by a number of video cameras. A diagram of the arrangement is given in *fig. 4*.

The following two models of visual search can be made. The pattern can be systematically scanned, e.g. 'line by line', or the searching movements of the eye may be completely arbitrary. In both cases the movement takes place in jumps, and during each fixation interval an area of the size of the conspicuity area is viewed in the search for the deviation. In a systematic search, i.e. line by line, the cumulative probability of

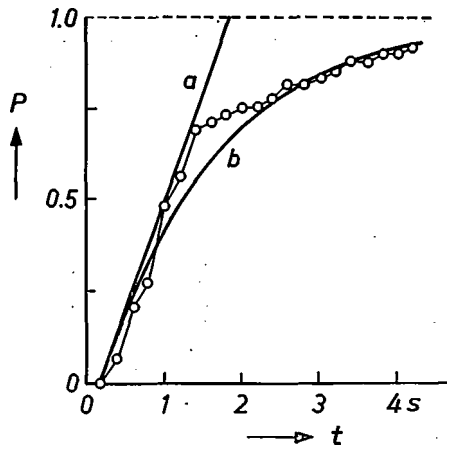


Fig. 5. The cumulative probability P of finding a deviating spot against the background of equally bright spots as a function of the search time t. Curve a gives the expected form for a 'systematic' visual search, curve b that for random searching movements of the eye. In the latter case it may take a very long time before the object sought is found. The points indicate measurement results. They show that against a structured background, as used in our experiments, the searching process can be described by assuming a non-systematic visual search. From the form of the curve drawn through a series of measured points the radius of the corresponding conspicuity area can be derived.

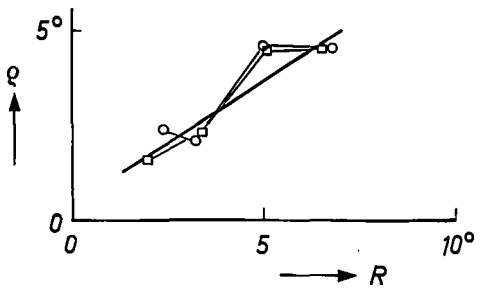


Fig. 6. Relation between the radius R of the conspicuity area as found in measurements with a single fixation, and the radius ϱ as calculated from the curve in fig. 5, giving the cumulative probability of finding a deviating spot as a function of time.

finding the deviation will increase linearly with time. In the random searching movement this probability will be an exponential function of time (*fig. 5*). In both cases the cumulative probability as a function of time will be directly connected with the size of the conspicuity area.

Fig. 5 also shows the results of a number of experiments that indicate that the exponential function is the best approximation to a visual search. The size of the conspicuity area as determined in measurements of this type shows a direct relation with the results of the determination mentioned above of the conspicuity area from a series of brief observations of the test pattern (*fig. 6*).

A later step in our experiments was to display test patterns containing two spots of a different diameter. The instruction here was to localize the large spot only (in fig. 1 the spot with the smaller conspicuity area). It was found that the observer directed his eyes towards the other spot, the one with the larger conspicuity area, more often than would correspond to arbitrary visual searching movements. The extent to which this happened was also found to be directly related to the size of the conspicuity area of the spot that was not searched for.

The concept of conspicuity area is thus found not only, as we have first shown, to be directly correlated with particular parameters of the test pattern, but also to be a good starting point for understanding visual search.

F. L. Engel

Ir F. L. Engel, formerly with the Institute for Perception Research, Eindhoven, is now with Philips Research Laboratories, Eindhoven.

A fast actuator for an anti-lock braking system

D. R. Skoyles

It is very important for road safety that cars should not skid when the brakes are applied suddenly. There is a danger of this on wet or icy roads, where the wheels can easily become locked. To prevent this from happening various anti-lock braking systems have been developed. This article describes an electro-hydraulic brake-pressure control with a number of new features. It was initially developed for testing electronic circuits for anti-lock systems and the design has given such good results in trials that it has been taken as the starting point for the further development of anti-lock braking systems.

Introduction

When the brakes of a vehicle are applied a certain amount of 'slip' must occur between the wheels and the road. The word 'slip' is used here as an indication that the braked wheels rotate more slowly than they would if they rolled freely [*]. When the brakes are first applied, the wheels continue to roll over the road, but the rate of rotation decreases because of the elastic deformation in the part of the tyre that forms the contact patch on the road surface. As the brake pressure increases, the speed of the wheels relative to the road decreases and eventually the wheel will lock. The degree of slip as the wheel approaches lock is a complicated function of the characteristics of the brake, the tyre and the road surface. Other factors such as the speed and deceleration of the vehicle also come into play.

When the wheels of a vehicle are locked, the stopping distance generally becomes larger, and if the front wheels are locked, steering becomes impossible. Most drivers are fully aware of this frightening loss of control and do not attempt to apply the brakes fully if the road surface appears to be slippery. Even experienced professional drivers are unable to brake at the maximum value as they cannot allow for the difference in characteristics from wheel to wheel.

In the interests of safety it is clear that wheel lock should be avoided. Some way must therefore be found of preventing the brakes from being applied too hard. If this can be done automatically, a driver would be able to brake fully in an emergency, while full steering control and normal handling will be maintained. Many practical solutions to this problem have been proposed: a few of them are now in use on a limited scale. This article describes a hydraulic pressure control or *actuator* for an anti-lock system; the device was devised and constructed for testing electronic control modules for such a system in a vehicle. At the time there was no existing actuator with a fast enough response for the purpose. The actuator was initially developed for private cars, but it may well be suitable for other vehicles. Many features of this actuator are used in more recent anti-lock systems under development.

Before we begin the actual descriptions of the design and construction of the actuator, we should first consider the behaviour of a skidding vehicle, tyre characteristics and the arrangement of anti-lock systems.

D. R. Skoyles, M.I.E.E., is with Mullard Research Laboratories, Redhill, Surrey, England.

Construction and use of anti-lock systems

There are three fundamental parts to any anti-lock system: a sensor that measures the wheel speed (or some related speed), a processor unit (typically electronic or fluidic) and a mechanical actuator that can briefly reduce the brake pressure and then restore it again as long as there is pressure in the master cylinder of the braking system. The wheel sensor may measure the speed of a single wheel or of a number of wheels. The output quantity is a signal whose frequency or voltage is proportional to the wheel speed. This signal is applied to the processing unit, where it is used to predict an imminent wheel lock. If necessary, the output signal from the processor is made to operate a solenoid valve or other actuator to reduce the brake pressure. More than one such actuator may be used. In most anti-lock systems the brake force falls while the solenoid (or other device) is energized, and increases when the solenoid is switched off (after the danger of a wheel lock has passed). Since the *passive* state of the pressure regulator corresponds to normal

braking, a certain measure of fail-safe behaviour is automatically built into the system.

Ideally the combination of sensor control circuit and actuator would be fitted to each wheel separately, allowing each tyre to have the optimum grip on the road. This would give the driver maximum braking and steering control, but it is costly. In many existing systems a single actuator is used to control one axle or one pair of wheels. In such cases the control must be applied to the rear wheels, as the following argument shows.

Although an anti-lock system will in general reduce the stopping distance, its most important function is to prevent loss of control of the vehicle. At first sight it would appear that it is more important to prevent the front wheels from locking, since these do the steering. (This would also appear to give a shorter stopping distance, since most of the braking force is provided by the front wheels owing to weight transfer.) This is not so, however: a moving vehicle with a single pair of wheels locked will eventually travel in the original direction but with the rolling pair of wheels trailing [1]. If only one pair of wheels is to be controlled, it must therefore be the rear wheels.

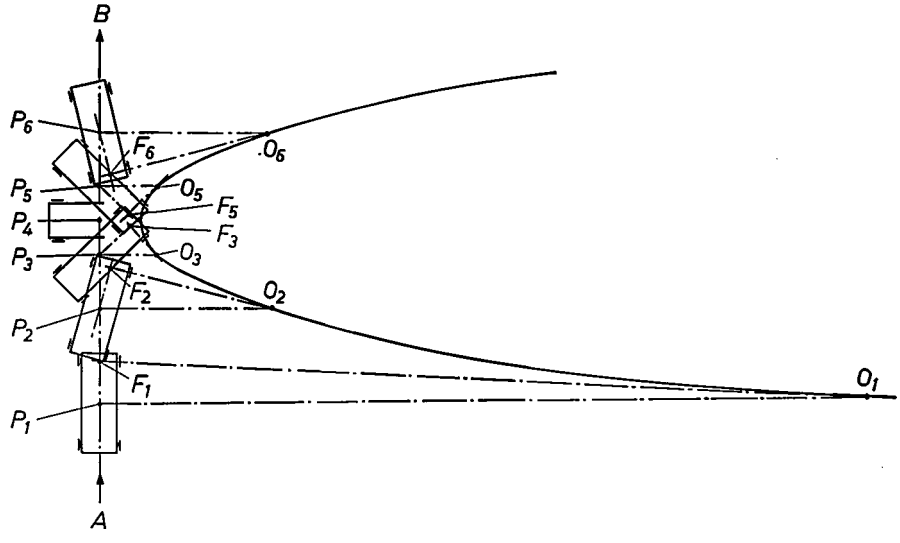


Fig. 1. Illustrating the behaviour of a vehicle if wheels become locked during braking.

The behaviour of a vehicle when wheels are locked during braking will be discussed with the aid of *fig. 1*. The vehicle is travelling along the line AB. It is assumed that only the *rear* wheels lock, at the instant when the centre of gravity of the vehicle is at the point P_1. If the long axis of the vehicle now deviates slightly from AB, the front of the vehicle starts to move in the direction determined by the position of the front wheels, while the centre of gravity of the vehicle continues to move in the direction AB and the locked rear wheels can skid equally in any direction. If the driver now keeps the front wheels pointing in a direction parallel to AB, the vehicle continues to move with its long axis parallel to AB until it comes to a stop. This motion is very unstable, however, and requires great skill on the part of the driver. If the driver does not alter the position of the front wheels, then small random deviations will cause the vehicle to move in a direction other than AB, e.g. at right angles to O_1F_1. The centre of gravity continues to move along AB, i.e. at right angles to O_1P_1. The vehicle therefore describes a circle, centred on O, and starts to rotate in a clockwise direction at an angular velocity directly proportional to the velocity in the forward direction and inversely proportional to the radius O_1P_1. After a short time the centre of gravity of the vehicle has moved to P_2; the deviation of the front with respect to AB has become larger and the centre of the rotation has moved to O_2. Since $O_2P_2 < O_1P_1$ the angular velocity of the rotation has increased. In this way the vehicle goes via position *3* to position *4*, where it makes an angle of 90° with the original direction. Here the front is stationary for an instant and the vehicle rotates about the centre of the front axle. In position *5* the rear wheels move in front and the centre of the rotation is at O_5, while the radius O_5P_5 now becomes greater and the angular velocity decreases. In position *6* the vehicle is almost in stable motion again, parallel to AB, but with the locked rear wheels in front. A number of second-order effects have been neglected in this qualitative description [2].

If anti-lock control is applied to the rear wheels only, the front wheels will lock and skid when the vehicle is strongly braked under adverse conditions. If the long axis of the vehicle now

[*] This is the technical definition of slip, which we shall use in this article. A sliding movement of a car with locked wheels will be called 'skid'.

[1] H. Darwin and C. V. Burton, Side-slip in motor cars, Engineering, Sept. 1904.
G. Jones, The skidding behaviour of motor vehicles, Proc. Auto. Div. Instn. Mech. Engrs 1962/63, No. 1.

[2] Fig. 1 and the associated explanation are mainly taken from: J. Bradley and S. A. Wood, Factors affecting the motion of a four-wheeled vehicle when some of its wheels are locked, Proc. Instn. Auto. Engrs 25, 59-62, 1930/31.

deviates slightly from the direction of travel, the frictional forces between the front tyres and the road *in the direction of motion* will produce a torque that will tend to swing the vehicle further round. However, the *transverse* forces on the rolling wheels (the rear wheels here) will set up a much larger torque, tending to restore the long axis of the vehicle to the direction of travel.

The tyre-road interface

Since rubber is elastic, the tyre and the tread have a certain flexibility, and this has an influence on the friction between tyre and road [3]. *Fig. 2* shows, greatly exaggerated, the deformation of the rubber where the wheel contacts the road during braking.

As a point on the tyre periphery moves into the contact patch its effective distance from the centre of the wheel decreases and it is at the same time subjected to a shear force which distorts it. The distortion is a function of both the shear (retarding) force and the weight acting on the wheel. For a given weight, there is an instant when the distortion with increased braking reaches a maximum value at which it is too great to be maintained by the frictional drag of the road surface and at this point sliding of the rubber occurs. Break-away and consequently sliding begin first at the rear edge of the contact patch where the distortion is greatest and the vertical contact force lowest. As the retarding force increases, the tread distortion increases and the break-away spreads towards the centre of the patch, where the vertical contact force is greatest. Eventually, as the brake pressure increases further, the shear force inevitably reaches a value which causes sliding even at the centre [4]. Before this point is reached the retarding force available is always increasing but once the sliding area has grown sufficiently to include the centre of the original contact patch, the force from the gripping rubber drops rapidly as sliding starts, lowering the friction force from the static to the sliding value. This decreases the effectiveness of the front part of the contact patch. The action is cumulative: increase of sliding area results in still lower restoring forces and the wheel rotation rapidly degenerates to the locked condition. This process is illustrated in *fig. 3* where vehicle retarding force and braking pressure are plotted against time.

The laws of friction impose upper limits on the maximum retardation forces available at the tyre-road interface. For a typical tyre on a good, dry level surface a vehicle retardation of about 0.8 g is possible. (1 g is the acceleration due to gravity.) This figure is independent of the tyre contact area and of the mass of the vehicle.

Fig. 4 shows how the angular retardation $\ddot{\theta}$ achieved on various surfaces depends on the slip [5]. The slip S is defined by:

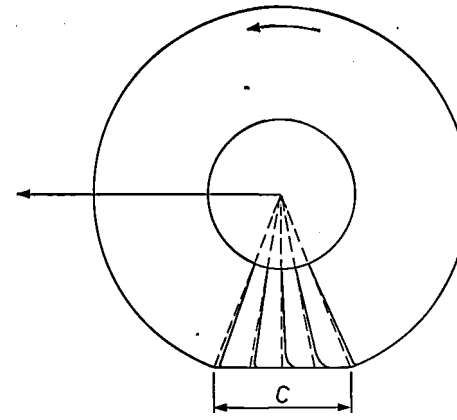


Fig. 2. Distortion of a tyre by the drag of the road when a wheel is braked. C represents the contact patch and the arrow represents the direction of motion of the vehicle. The dashed lines represent radial planes through the centre of the wheel; the solid lines show schematically how these planes are distorted when the wheel is braked. Part of the distortion lies in the side wall of the tyre and part in the tread. The distortion is greatest near the rear of the contact patch.

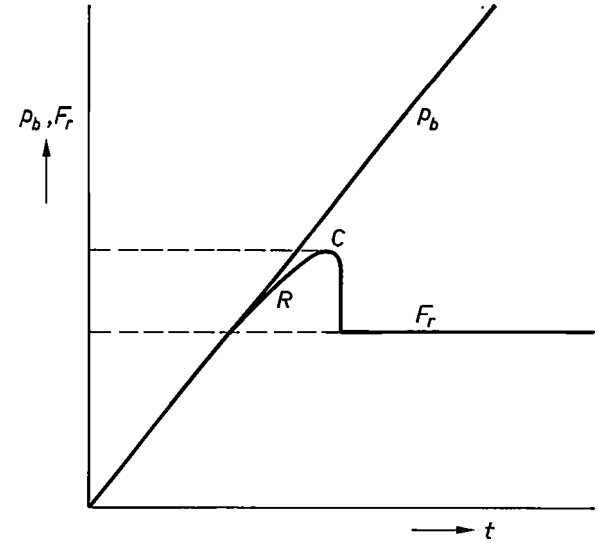


Fig. 3. Variations of the frictional force between tyre and road surface for increasing brake pressure. The frictional force F_r and the brake pressure p_b are plotted in arbitrary units. At R there is still complete grip of the road surface; C indicates the point where the break-away from the road surface has expanded to include the centre of the contact patch. The wheel then locks and F_r decreases to the constant value that applies for the sliding tyre.

$$S = \left(1 - \frac{\dot{\theta}_w}{\dot{\theta}_v}\right) \times 100\%,$$

where $\dot{\theta}_w$ is the actual angular velocity of the wheel and $\dot{\theta}_v$ is the angular velocity the wheel would have if rolling freely.

Although other measurements could give results somewhat different from those of fig. 4, it is in general true that if wheel slip could be held between 12 and 15%, a retarding force would be available which (on most surfaces) would be higher than the retarding force available from a locked wheel.

The retardation-slip characteristics show that even light braking (low retardation) is accompanied by a certain amount of wheel slip. This does not necessarily imply sliding as such, but each part of the rubber tread is distorted as it forms a part of the contact patch during braking and each new distortion occurs at the expense of the angular velocity of the wheel. This produces a rotation which is slower than that of a freely rolling wheel.

The higher the peak retardation ratio (= peak retardation divided by retardation at $S = 100\%$, see fig. 4), the greater the increase in stopping power given by a good anti-lock system. On wet slippery surfaces the peak retardation ratio of the slip characteristic (fig. 4) depends on the ability of the tyre tread to break through or squeeze out the moisture film between rubber and road. Since this is more difficult at high speeds, the peak of the slip characteristic has a value which decreases with speed. It is also very dependent on tyre-tread design and surface texture. A worn tyre with a tread of less than 1-2 mm has inadequate drainage [6] and the braking force on a smooth wet surface is therefore less than that with a full tread. A rough road surface allows improved drainage and the difference between a worn and an unworn tread is less marked. Experiments have shown that tread is unimportant on normal dry surfaces, where even bald tyres give excellent performance.

A good, dry road surface offers higher road adhesion and higher braking torques are possible without the danger of wheel lock. *Fig. 5* shows some slip characteristics for various road surfaces. The retarding torque given by the road-tyre drag is plotted (instead of the retardation) as a function of the slip S.

When the high-slip region of the torque-slip curve is nearly horizontal, the retarding force of the locked wheel is not greatly different from the peak braking torque near 15% slip. In this situation the reduction in brake pressure that the anti-lock device must provide to allow the wheel to recover to the stable side of the curve must in fact be less than in the case of a steeper slip characteristic. The way in which the anti-lock device reduces the brake pressure will be discussed in the next section.

Gravel surfaces and loose snow behave anomalously (see figs. 4 and 5) because the material piles up in front of the wheels.

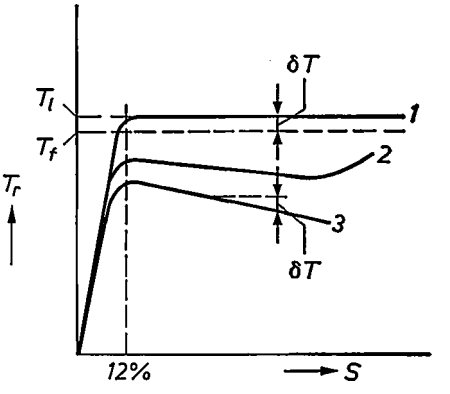


Fig. 5. Braking torque T_r available on various surfaces as function of the slip S. *1* good surface, *2* gravel (the hump is due to piling up). If T_l is the braking torque at which the wheel started to lock on a good surface, a small reduction δT to T_f will return the wheel to a safe braking condition. The same torque reduction would be inadequate for a more slippery surface, as can be seen from curve *3*.

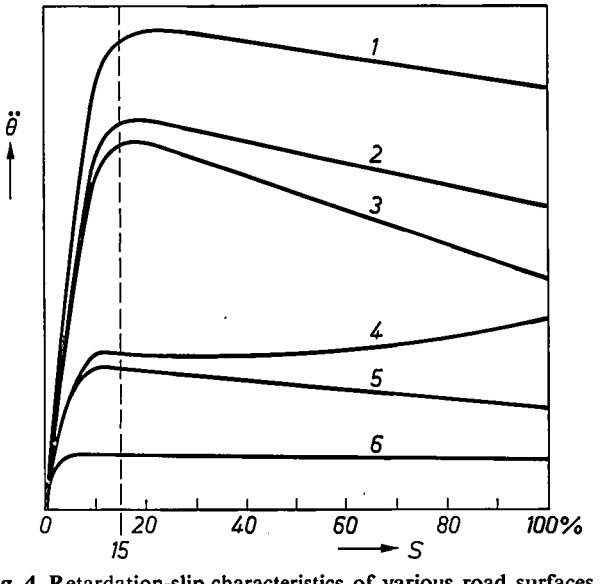


Fig. 4. Retardation-slip characteristics of various road surfaces, showing that the peak retardation is mostly attained for a slip of between 10 and 15%. *1* dry asphalt, *2* wet asphalt (thin water film), *3* wet asphalt (thick water film), *4* fresh snow, *5* packed snow, *6* ice.

Principles of anti-lock systems

Before discussing the principles on which an anti-lock system can be designed, it is important to recognize the surprising fact that one essential piece of information — the true road speed of the vehicle — is not readily available. As we saw earlier, the peripheral speed of the wheel is only equal to the road speed if there is zero slip. As soon as slip appears, the two speeds differ. The wheel speed can easily be measured by a transducer at the wheel. To determine the slip, however, the actual speed relative to the road must be known. Apart from such obvious methods of measurement as

[3] G. Temple, The dynamics of the pneumatic tyre, Endeavour 15, 200-205, 1956.
[4] K. N. Chandler, Theoretical studies in braking, Part I: Effects of longitudinal slip for a single wheel, Proc. Auto. Div. Instn. Mech. Engrs 1960/61, No. 4.
[5] P. Müller and A. Czinczel, FISITA 14th Congress, London 1972, p. 3/92.
[6] G. C. Staughton, The effect of tread pattern depth on skidding resistance, Report of the Transport and Road Research Laboratory LR 323, 1970.

a fifth wheel (not acceptable for private cars), or more complex methods such as Doppler radar (too expensive), there is no simple way of doing this. The road-speed information necessary for anti-lock control — or some approximation to it — must therefore be extracted in another way.

An *estimate* of the *minimum* road speed at any given instant can be derived from the wheel speed at the moment the brake is applied and an assumed maximum attainable deceleration of the vehicle. This can be conveniently obtained electronically by using a voltage proportional to the wheel speed and a circuit in which the maximum deceleration is simulated. This circuit gives a reference voltage proportional to the estimated road speed. The voltage corresponding to the *instantaneous* wheel speed is now compared to the above reference voltage. If it is less than 85% of this value this means that the slip of the wheel is excessive.

The best possible braking would be obtained if the brake pressure could be always held at the value coinciding with the maximum retarding force available at each instant from the road (see figs. 4 and 5). If the wheel is held at this 'state of maximum adhesion' the maximum lateral or steering force is in general obtained [7]. In practice the optimum brake pressure cannot be used because regulating the pressure to this (changing) optimum value requires the true slip of the wheel to be known accurately at every instant. It might appear that a good compromise would be to regulate the brake pressure at a level lying at a more stable point *just below* the peak adhesion region of the retardation-slip curve (fig. 4). However, this also turns out to be impracticable because of the variation in road-surface friction, and there are also problems of long-term stability. In practice the pressure is therefore 'modulated' about the optimum, i.e. cycled about the point of maximum adhesion, high slip values causing a reduction of mean pressure. This allows for varying surfaces and changes of weight between the wheels. Since lower slip values arise repeatedly, a fairly simple method can be used for supplying the electronic control module with regularly updated computed information about the wheel slip: this information is essential for determining the correct instants for the relaxation and re-application of the brake pressure.

For a level-force-slip characteristic (fig. 5, curve *1*), reduction to below the locking pressure may appear to provide a braking force lower than that of a locked wheel but this is not necessarily so. The pressure may be modulated about a value well below the optimum pressure, without any reduction in the average vehicle-retarding force. Wheel inertia plays an important part here. The following statement may help to clarify this: if the vehicle is moving fast with all wheels locked owing to excessive brake pressure and the brake pressure is instantaneously reduced to zero, full 'braking' is maintained until the wheels have been accelerated (by the road) to about 85% of the vehicle speed (i.e. about 15% slip). The retarding force after zero brake pressure is produced here not by brake torque but by wheel inertia.

The sum of the torques acting on the wheel is zero:

$$T_r + T_b + I\ddot{\theta}_w = 0,$$

where T_r is the road torque, T_b is the brake torque, I is the moment of inertia of the wheel and $\ddot{\theta}_w$ is the angular acceleration. When the brakes are applied, the stored energy of the vehicle is reduced but so is the stored energy of the wheels as they decrease in speed. When the brakes are released the acceleration of the wheels by the road is associated with an inertial reaction torque

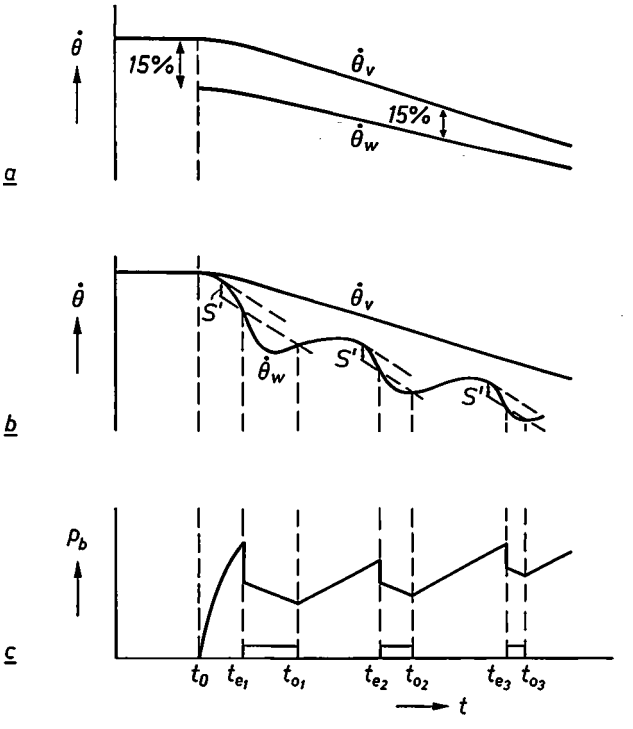


Fig. 6. *a*) The angular velocity $\dot{\theta}$ of the wheel, plotted as a function of time. $\dot{\theta}_v$ represents the vehicle road speed expressed in terms of the angular velocity of a (hypothetical) freely rotating wheel. For $t < t_0$ the speed is constant; at $t = t_0$ the brakes are applied. The deceleration $\ddot{\theta}_v$ of the vehicle, which is given by the slope of the curve $\dot{\theta}_v$, then changes from $\ddot{\theta}_v = 0$ to $\ddot{\theta}_v = -K$, where K is assumed constant. The maximum value of K in practice is found to be about 0.8 g. When the brakes are applied there is always some slip. The curve $\dot{\theta}_w$ shows how the wheel speed would vary in the ideal case, i.e. braking such that the slip remains between 10 and 15% (see the peak retardation in fig. 4). *b*) Variations of the true wheel speed $\dot{\theta}_w$. The upper dashed line is the tangent to the curve $\dot{\theta}_w$ at a point where the slope $\ddot{\theta}_w$ corresponds to a deceleration of 1.5 g. This line is the velocity reference. The lower line is drawn parallel to the upper line and below it such that the distance S' represents a 5% slip with respect to the velocity references. This implies that at the time t_{o_1} the deceleration of the wheel is well in excess of the reference at 1.5 g; at this point the solenoid valve is energized. *c*) The variation of the brake pressure p_b with time, as controlled by the energization of the solenoid valve. (The shape of this curve is discussed later in this article, with the construction of the valve.) At t_{o_1} the solenoid valve is de-energized. t_{e_2}, t_{o_2} etc. are subsequent cycles of operation.

$I\dot{\theta}_w$ which has the same sense as T_b. The energy required to build up this torque is extracted directly from the kinetic energy of the vehicle.

Many anti-lock control devices make use of the wheel inertia effect; it is a feature which greatly facilitates the realization of practical systems.

When the control circuit detects a potential wheel lock, a solenoid valve is opened to relieve the brake pressure. After a certain inevitable delay, the grip of the brake on the wheel will have decreased sufficiently for the drag force of the road to re-accelerate the wheel: the slip therefore starts to decrease again. When the slip has been sufficiently reduced, the brake pressure can be increased again by closing the solenoid valve. The whole process — detection of excessive slip, opening of solenoid valve and closing it — is now repeated as shown schematically in *fig. 6a, b* and *c*. The brake pressure and the slip are thus continually cycled about the point of maximum adhesion (fig. 4).

The *first* energization of the solenoid valve takes place just after the wheel has been decelerating at a rate greater than that corresponding to a vehicle deceleration of 1.5 *g* (well above the maximum vehicle deceleration attainable). Actual switch-on occurs when the wheel slip amounts to more than 5% relative to the reference speed corresponding to a deceleration of 1.5 *g* (hence at time t_{e_1} in fig. 6*b*). This is not the ideal instant for the start of the energization, but the actual slip will be greater than 5% since vehicle deceleration will always be far lower than 1.5 *g*, and the vehicle speed will always be above the computed reference.

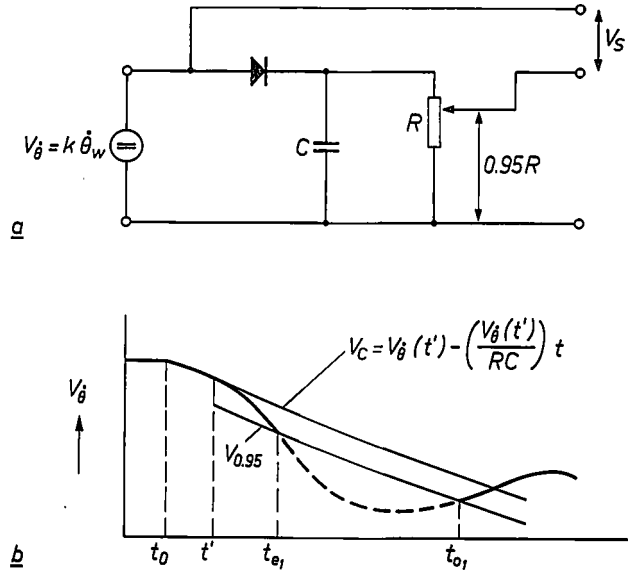


Fig. 7. *a*) Elementary circuit for the comparison of the wheel-speed sensor voltage with an artificial reference corresponding to a wheel deceleration of 1.5 *g*. The output voltage $V_{\dot\theta}$ from the wheel sensor (a voltage generator) is proportional to the wheel speed, $V_{\dot\theta} = k\dot\theta_w$. The capacitor is charged to this voltage via the diode but the charge leaks away at a rate determined by *CR*. *R* and *C* are chosen such that the charging rate corresponds to a deceleration of 1.5 *g*. *b*) Variations of the voltage $V_{\dot\theta}$ from the wheel sensor. As soon as the wheel deceleration is lower than 1.5 *g* the voltage V_C across *C* is equal to $V_{\dot\theta}$, but if the deceleration exceeds 1.5 *g* (at time *t'*), $V_{\dot\theta}$ falls below V_C. A sensing circuit gives a signal if the voltage V_S becomes negative, i.e. if $V_{\dot\theta}$ falls below 0.95 V_C (the line $V_{0.95}$). This signal is used to energize the anti-lock solenoid (at the time t_{e_1}). The slip is then about 5% greater than it was when the reference was first exceeded. The energization is removed again at the time t_{o_1} when $V_{\dot\theta}$ again becomes larger than $V_{0.95}$.

The slope of the *ideal* curve $\dot\theta_w$ in fig. 6*a* corresponds to a wheel-speed control such that there is always 15% of slip, i.e. such that $\dot\theta_w$ remains continuously 15% below the curve $\dot\theta_v$. It might appear to be better to wait until the wheel speed had dropped to the value corresponding to 15% slip (maximum adhesion) before energizing the solenoid to relieve the brake pressure, but in practice it is more advantageous to energize as soon as possible in the case of rapid brake application (slow application allows the desired slip value of 15% to be reached). In addition, it is more convenient to operate with a single slip setting for the wheel sensor (the tapping in fig. 7*a*, see below). When the slip of the wheel begins to decrease (re-acceleration to the vehicle speed) the brake pressure has to be re-introduced sufficiently slowly to enable the wheel to reach the stable side of the curve to the left of the peak in fig. 4). A repetitive pressure-cycling action is then possible. If the stable side of the curve were not reached, the control circuit would not get enough information to assess the slip and the wheel would lock.

Fig. 7a gives a highly simplified example of a circuit in which the artificial reference voltage corresponding to a vehicle deceleration of 1.5 *g* is compared with the voltage from the wheel-speed sensor. This voltage $V_{\dot\theta}$, which is always proportional to the wheel speed $\dot\theta_w$, is applied to the capacitor *C* via a diode. A resistance *R* is connected across the capacitor, giving a constant *RC* corresponding to a deceleration of 1.5 *g* (*RC* is usually greater than the cycle time $t_{e_2} - t_{e_1}$):

$$V = V_0 e^{-t/CR} \approx V_0 \left(1 - \frac{t}{RC}\right) = V_0 - \frac{V_0}{RC} t.$$

The voltage across *R* cannot therefore fall faster than $\dot V = V_0/RC$; see the line V_C in fig. 7*b*.

For a given wheel speed $\dot\theta_w$, a voltage $V_{\dot\theta} = k\dot\theta_w$ is generated. The capacitor is charged to this voltage, which therefore appears across *R*. On sudden braking, the wheel velocity can fall faster than the fixed rate $\dot V = V_0/RC$, which is made to correspond to a vehicle deceleration of about 1.5 *g*. Whenever this happens we know that $\ddot\theta_w$ has a value larger than that corresponding to 1.5 *g*. The solenoid valve is energized if the voltage $V_{\dot\theta}$ falls below the voltage at a tapping on *R* corresponding to 0.95 *R* (at the time t_{e_1}, if the curves $V_{\dot\theta}$ and $V_{0.95}$ intersect). This will occur at a slip value 5% in excess of the reference line. When the speed of the wheel has increased again such that $V_{\dot\theta} > V_{0.95}$ (at time t_{o_1}), the solenoid valve is de-energized, and the

[7] K. E. Holmes and R. D. Stone, Tyre forces as functions of cornering and slip on wet road surfaces, Report of the Transport and Road Research Laboratory LR 254, 1969.

whole cycle is repeated again if required. The circuits used in practice are more complex to allow for a constant slope at all wheel speeds and also to accommodate all kinds of delays and second-order effects.

An actuator for an anti-lock system

The actuator used has the form shown in *fig. 8*. It consists of a solenoid valve, a variable restrictor and a pump. The pump is driven directly by the wheel.

The actuator was designed with the seven following aims in view:

It should be a small compact unit which can be housed within the wheels (no vulnerable pipes or signal wires across the vehicle, potential sources of failure).

The actuator should not affect the normal operation of the brakes.

The instant at which the brakes are reapplied should depend on the decrease in slip during an anti-lock period and also on anti-lock behaviour during previous cycles.

The brake pressure should not be allowed to fall so far that under-braking results when the wheel is freed from lock.

Fail-safe arrangements should be included in case of circuit or wiring failures or other faults.

There should be a single solenoid valve for brake-pressure control.

The control system should be completely independent of the driver.

Incipient wheel lock causes the control circuit to open the solenoid valves and hence reduce the brake pressure. When the wheel has recovered to a safe condition (left-of-peak in fig. 4) the solenoid is closed and the fluid that has flowed from the brake cylinder B is returned to the pressurized master cylinder M. If this were not done the drivers' foot would continue to move towards the floor in an effort to maintain the pressure in a 'leaking' system. At the same time the pressure in the brake cylinder increases again. To obtain the maximum braking this should happen as quickly as possible; the rate of pressure build-up should not be affected by factors other than the behaviour of the wheel.

The variable restrictor

On a good road surface the wheel can recover speed rapidly because of the high frictional drag of the road on the tyre. For this reason a higher rate of pressure recovery is required on a good surface than on a poor surface. This is arranged by designing the actuator in such a way that the pressure-recovery rate is a function of the total time for which the solenoid has been energized. This is done by the variable restrictor V (fig. 8). The longer the solenoid is opened (poorer surface) the more the cylinder D with the plunger J_1 is driven to the right and the slower the build-up of pressure in the brake after the solenoid has closed again. On a good road surface the wheel recovers its speed very rapidly after a high slip: the solenoid is only energized for a short time and the plunger hardly

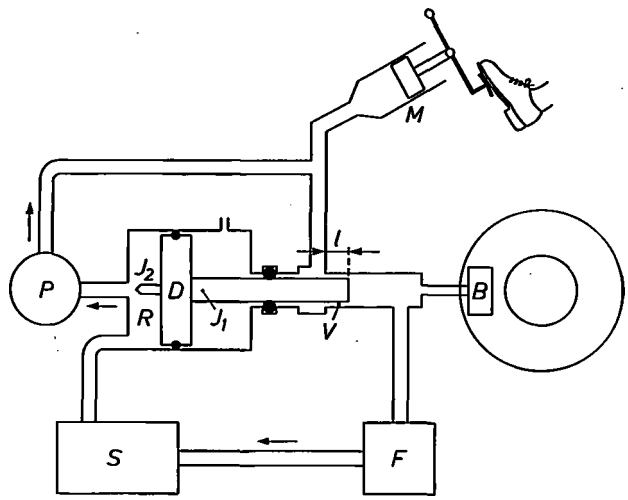


Fig. 8. The anti-lock braking system. The actual braking system consists of the master cylinder M and the brake cylinder B that operates the brake of a wheel. The main parts are a solenoid valve S which relieves the brake pressure when the wheel is about to lock, a variable annular restrictor V with plunger J_1, which determines the rate of pressure build-up in the brake after the solenoid valve has closed again, and a pump P which returns diverted brake fluid to the master cylinder. The pump can be located in the wheel and driven by it. F is a fail-safe control valve which closes in the event of faulty operation of S. R is a reservoir for the diverted brake fluid; this brake fluid displaces the piston D, so that the quantity of fluid in R determines the amount of closure of the restrictor V (see also fig. 9). J_2 is a second plunger mounted on D, which modifies the rate of brake-pressure recovery.

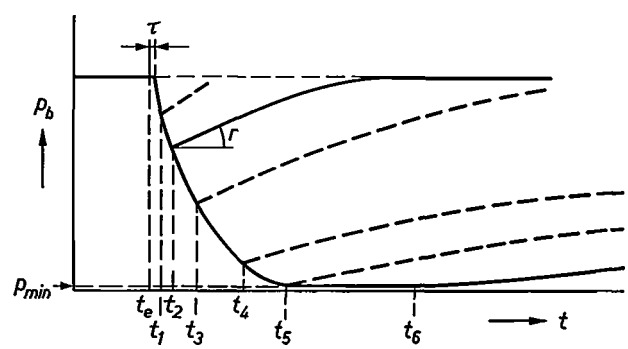


Fig. 9. Pressure in the brake cylinder decreasing as a function of time, and its dependence on the solenoid energization time. t_e represents the instant at which the solenoid is energized. There is first a slight delay τ before the pressure begins to drop (a few milliseconds). If the energization continues for a long time the pressure can drop to p_{min}. A typical energization lasting for a time $(t_2 - t_e)$ allows a build-up of pressure at the rate indicated by the angle r after the solenoid has closed. A shorter energization results in a more rapid pressure build-up. Longer energizations are accompanied by a slower increase until t_6 is exceeded. At this time the reservoir R (fig. 8) is full and pressure would be restored, in the absence of pump P, even if the solenoid remained energized. If, however, the pump was operating, all fluid would be returned to the master cylinder and no pressure build-up would be possible until the solenoid was de-energized.

penetrates the opening, so that the brake pressure can build up sufficiently rapidly to keep up with the rapid recovery of wheel speed. This is illustrated in *fig. 9*.

The variable restrictor cannot be designed to provide the optimum rate of pressure rise for all kinds of road surfaces, because its restrictive effect depends on factors other than the position of D (fig. 8), which are determined by the amount of brake fluid in the reservoir R; the pressure rate also depends on the viscosity, for example. However, the variable restrictor allows high rates of pressure recovery for good surfaces without compromising the performance on a poor surface. Also, because the closure of the restrictor depends on the position of D in the reservoir, the control becomes self-corrective, i.e. if the wheel under anti-lock control on a poor surface suddenly encounters a good surface, the restrictor is rapidly withdrawn since the fluid is quickly pumped from the reservoir and there is no further replenishment via the solenoid. The restrictor adapts rapidly to the new conditions and allows a higher rate of pressure recovery.

Control of the brake pressure in the manner described relies on the fact that the slip decreases to a lower value than the optimum (15%, see fig. 4). Any system in which the vehicle speed is not actually measured or computed has to rely on this. If for any reason the slip does not decrease from the original high value to a value in the stable region, the control system will measure too high a value of slip. However, under these circumstances the variable restrictor extends the control logic in the sense that each solenoid energization ensures a succesively lower rate of pressure build-up (more and more closure of the restrictor) until the slip does become sufficiently small. Although in this method of slip reduction the slip temporarily exceeds the optimum value, this is preferable to a complete lock of the wheel. The variable rate of pressure build-up is very important and eases the task of the electronic control unit.

The restrictor plunger J_1 is completely withdrawn for normal braking. When fluid has been diverted from the brake via the solenoid valve to the reservoir (R, fig. 8), the brake pressure exercises pressure on the fluid in the reservoir via the plunger. When the reservoir is nearly empty, the outlet to the pump is partially blocked by the plunger J_2. At this point the plunger J_1 takes up a position such that when anti-lock control has ceased the brake fluid can flow into the brake cylinder at the maximum rate for good surfaces. The rate at which this happens increases as the reservoir finally empties, despite the damping action of the plunger J_2. The maximum time for the wheel to regain speed during an anti-lock cycle is in the region of 0.1 s; this is short compared with the time taken for J_1 to reach the open position when anti-lock control has ceased.

The rate of increase of the brake pressure after an anti-lock energization of the solenoid valve is controlled mainly by the restrictor but it also depends on the viscosity of the brake fluid, the pressure in the master cylinder, the brake pressure, the capacity of the brake cylinder, the compression of the brake linings, the spring pressure of the brake calipers, the cross-section of the plunger and its radial clearance (referred to above qualitatively as the *closure*). The flow rate of the brake fluid through the restrictor is given by

$$\dot{q} = \frac{\pi d g^3}{12 \eta l} \Delta p,$$

where $\dot{q}$ is the volume flow in m^3/s, d the diameter of the restrictor in m, g the annular gap in m, Δp the pressure difference in N/m^2, l the length of the restrictor in m and η the viscosity in Ns/m^2. The formula shows that the flow rate is very sensitive to changes in the clearance of the restrictor plunger. Very close tolerances are therefore required.

The viscosity of brake fluid at 20 °C is ten times the viscosity at 100 °C and a tenth of the viscosity at −20 °C. Changes in viscosity can be compensated by using materials with different expansion coefficients for the restrictor housing and plunger; this gives an annular clearance which varies linearly with temperature. The change in flow rate is then proportional to the square of the temperature but the variation in viscosity with temperature is logarithmic. The compensation can therefore never be perfect and further compensation may be necessary for optimum operation at extreme temperatures.

The pump

The pump (P in fig. 8), driven directly by the wheel, must have the following characteristics:

a) Even at low road speeds it must be able to return all the fluid to the master cylinder, i.e. it must be able to rapidly empty the reservoir when the solenoid valve is closed.

b) At high speeds and low brake pressures the pumping efficiency should fall to a low value.

The first feature ensures that the anti-lock device operates well even at low speeds. The second feature ensures that the wheels have sufficient time to regain speed even after locking at high speed and low brake pressure. This feature is fortunately present in most simple pumps. In a typical case a simple plunger pump may be used, the plunger being pushed against a shallow cam (which revolves with the wheel of the vehicle) by the brake fluid in the reservoir (R in fig. 8). At low brake pressures the pump plunger cannot move rapidly because the passage between the reservoir and plunger acts as a restrictor. Also, at high speeds the pump plunger will be unable to follow the cam completely. The rate of pumping $\dot{Q}$ is therefore controlled by the restriction to the pump inlet. This is shown in *fig. 10*.

The effect of the pumping action is to reduce the amount of fluid in the reservoir R and hence give a displacement to the left of the plunger J_1 of the variable restrictor V. The rate of flow of fluid from master cylinder M to the brake cylinder B therefore increases

as pumping continues and this enables the system to adapt to a good surface after having been on a poor surface. High vehicle speed would theoretically give scavenging (emptying) of R at too high a rate (the scavenging rate is proportional to the speed) but the restrictive effect of J_1 and the connection between R and P (fig. 8) is sufficient to limit this to a suitable value without the necessity for any additional impedance.

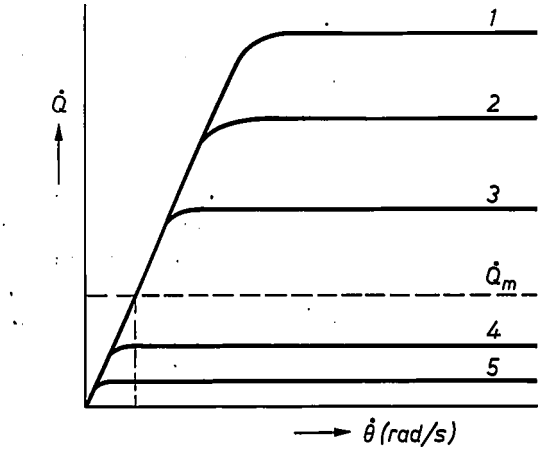


Fig. 10. Illustrating the pumping rate $\dot{Q}$ as a function of wheel speed $\dot{\theta}$, at various positions of the plunger J_2 (fig. 8). Curves *1, 2, 3* are for high and medium brake pressures, showing the pumping rates $\dot{Q}$ attainable when the plunger J_2 permits full flow. The line Q_m represents the maximum pumping rate when the plunger does limit flow. Curves *4* and *5* represent the rates for low brake pressures.

The solenoid valve

When not energized the solenoid valve is closed. At pressures up to about 20 000 kN/m² (3000 p.s.i., 200 atm) it completely prevents the flow of fluid. When energized it must release fluid at a rate which permits a brake-torque decay rate of about 2100 Nm/s and, equally important, when de-energized it must interrupt and stop the flow of fluid very rapidly. The operating current for the solenoid has to be provided by the car battery and in practice this means that the maximum current is about 10 A. The armature pull-in time should be below 7 ms at this current and any reduction of this time improves the braking performance.

The anti-lock system described here uses a solenoid with a flow control which allows an *initial* rapid drop of fluid pressure corresponding to a decrease in brake torque of 180 Nm, followed by a *linear* fall rate corresponding to 14 000 Nm/s (matched to the brakes of the test vehicle). If the initial drop in pressure alone determined the 'modulation depth' (the relative decrease in brake pressure, see fig. 6*c*) this would give rise to ± 17% of modulation on a dry surface (coefficient of friction $\mu \approx 1$) and ± 100% modulation on a slippery surface ($\mu \approx 0.15$). The large modulation depth is necessary on slippery surfaces for the wheel to recover its speed with a reasonable acceleration (about 4 g). The maximum theoretical acceleration is about 7 g but only about 4 g is usually available on a slippery surface, because there is always some residual brake pressure due to seal resilience. The fixed initial drop in pressure at t_{e_1} in fig. 6*c* provides an effective and simple way of obtaining the smaller pressure modulation required for good road surfaces. It would be difficult to control a fast pressure drop accurately by electronic methods, on account of the delay that would occur between the signal from the wheel sensor and the reaction from the actuator.

The linear drop in pressure (t_{e_1} to t_{o_1} in fig. 6*c*) is introduced partly to reduce the effect of transducer and circuit delays but mainly to reduce the turn-off delay of the solenoid itself.

The operation is as follows. In the much simplified diagram of the solenoid valve shown in *fig. 11* it can be seen that the piston G around the armature contains a number of orifices H. The pressure drop across an orifice is proportional to the flow, and this 'flow' force tends to close the valve at E, while armature current I will tend to open it. A constant flow results, but this can be altered by change in I. Since the valve at E is always partially closed, if I drops sharply to zero, the flow is quickly stopped.

The action described above comes from the interaction of various forces. With the valve opened there is an equilibrium of forces if

$$p_3A_3 - (p_1 - p_2)A_2 - F_S + F_A = 0,$$

where F_S represents the sum of the spring forces and F_A the magnetic force exerted on the armature; the other quantities are shown in fig. 11. Neglecting the small terms p_3A_3 and F_S, we find that the armature force F_A required is

$$F_A = (p_1 - p_2)A_2.$$

The force exerted on the armature in a coil is proportional to the square of the current I, i.e.

$$F_A = K_1I^2.$$

The pressure difference between the two sides of the piston G is proportional to the rate of flow of brake fluid through the orifice H:

$$p_1 - p_2 = K_2\dot{q}.$$

It is this flow that provides the linear drop in pressure referred to above. The equilibrium equation for the forces is then

$$K_1I^2 = K_2A_2\dot{q},$$

showing that the flow rate $\dot{q}$ is controlled only by the current, since all the other factors are constants.

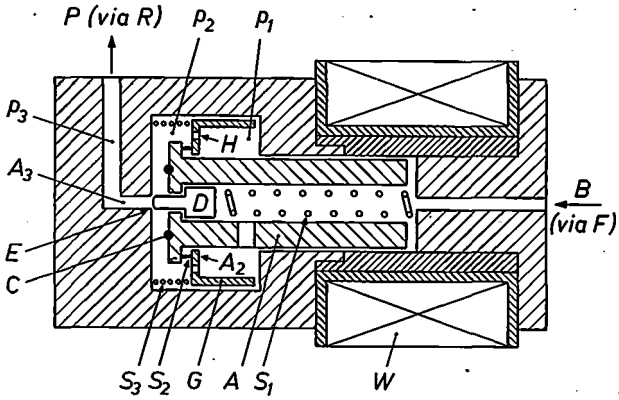


Fig. 11. The solenoid valve (*S* in fig. 8. Fluid enters from the brake cylinder *B* via valve *F*). *W* solenoid windings, *A* mild-steel armature. *E* main flow orifice. *G* concentric piston containing orifice *H*. *C* O-ring leak protection. S_1, S_2 and S_3 compression springs. p_1 is the pressure to the right of piston *G*, p_2 the pressure to the left of *G*, p_3 the pressure in the exit line to the reservoir *R* and pump *P*. The effective area of *G* is A_2, and the area of the exit line is A_3.

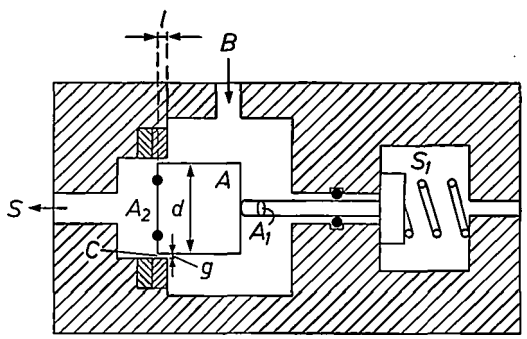


Fig. 12. Fail-safe flow-control valve (*F* in fig. 8). Brake fluid enters at *B* from the brake and leaves at *S* for the solenoid. During the first moments of braking before anti-lock functioning of the solenoid can begin any flow through the fail-safe valve must be due to faulty opening of the solenoid valve. The pressure drop produced across *A* by the flow drives the piston to the left, blocking further fluid flow. If, however, there is no fault, *A* must be cleared from the bore before the expiry of a certain time which depends on the brake pressure (see text and fig. 13). A_1 cross-section of piston rod. A_2 area of piston, *g* width of annular gap, *d* diameter of *A*, *l* length of channel *C* with piston in equilibruim position (zero brake pressure), determined by spring S_1.

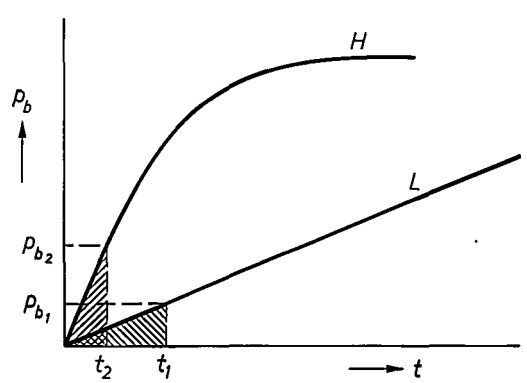


Fig. 13. Brake pressure p_b as a function of time *t* for hard braking *H* and light braking *L*. The shaded areas are equal. Their magnitude determines the time within which the wheel has decelerated so far that the solenoid valve is energized (the worst case is when the road is so slippery that the brake pressure only has to overcome the inertia of the wheel). The solenoid valve should therefore be tested within this time for correct operation, and any remedial action taken. If no fault is found the fail-safe valve must be opened before t_2 (for hard braking) or t_1 (for light braking).

Safety features

As mentioned earlier it is imperative that the anti-lock equipment should not increase the chance of the brakes failing in normal operation. Two fail-safe protections are therefore included: the first allows for the condition where there is a permanent leak through the solenoid valve, and the second takes care of the dangerous situation which would arise if the solenoid for some reason became permanently energized. The undesired consequences of a leak are easily avoided by including an O-ring seal in the sealing face of the armature of the solenoid valve (*C* in fig. 11). A light spring S_2 biases the armature to the left so that the O-ring makes good contact with the face. If there is a leak through the seal *E* the pressure inside the space sealed off by the O-ring is low and the brake pressure acting on the back of the armature presses the armature to the left with a force that exceeds the force available from the solenoid. This of course inhibits any anti-lock control in the event of emergency braking but it does allow normal braking in spite of solenoid leakage. When there is no leak, the armature is free to move to the right when the solenoid is energized.

The second and most important protection is provided by a hydraulic flow-control valve, illustrated schematically in *fig. 12* (*F* in fig. 8). It would be no use relying on electrical protection (switching off the power), as this would not be effective in the case of some types of short-circuit or in the event of a faulty seal or a 'sticking' armature in the solenoid valve.

The hydraulic control valve of fig. 12 detects fluid flow due to faulty operation or failure of the solenoid valve. Only under one special set of circumstances can flow due to this cause be distinguished from flow due to normal anti-lock operation. This situation arises at the onset of braking, before it is possible even theoretically for the anti-lock mechanism to have been called into use. Wheel inertia plays an essential role here, since a certain brake pressure will always be required to reduce wheel speed even if the wheel is not in contact with the ground. This means that a certain value of the pressure-time integral (*fig. 13*) must have been reached before the anti-lock mechanism can be called into operation. During this stage of inertial braking, therefore, any flow of fluid through the solenoid valve must be due to a fault. This flow is used to drive the piston *A* (fig. 12) to the left in order to stop any further flow.

If the system has not failed, the assembly must be driven to the right *before* the pressure-time integral has exceeded a given value in order that *A* may be cleared from the bore so that the solenoid can perform its anti-lock function if required. The piston is driven to

the right when the brake pressure acting on A overcomes the spring force.

When there is zero leakage at the solenoid valve, the net force on the piston A driving it to the right is $p_b A_1 - F_S$. This is approximately equal to $p_b A_1$ if the spring pressure F_S is low.

Movement of A is opposed by the viscous flow $\dot{q}$ of fluid through the annular channel C:

$$\dot{q} = \frac{\pi g^3 d}{12\eta x} \Delta p,$$

where d is the diameter of the piston, g the gap width, and x is the remaining length of the channel (the initial equilibrium length, when there is no brake pressure, is l). The pressure difference Δp across the channel is equal to the excess hydraulic pressure set up in the fluid to the right of the piston by the force acting on the left-hand face, i.e.

$$\Delta p = (p_b A_1 - F_S)/(A_2 - A_1).$$

Since the spring force $F_S \ll p_b A_1$ and $A_1 \ll A_2$:

$$\Delta p \approx p_b A_1 / A_2.$$

The velocity of the piston is then

$$\dot{x} = \frac{\dot{q}}{A_2 - A_1} \approx \frac{\dot{q}}{A_2} = \frac{\pi g^3 d}{12\eta x} \cdot \frac{p_b A_1}{A_2^2} = \frac{\pi^{1/2} g^3 p_b A_1}{6\eta x A_2^{3/2}},$$

where d is expressed in terms of A_2.

The time τ for A to clear the bore C, i.e. for the fail-safe valve to be put out of action, is thus given by the equation

$$-\int_l^0 x\,dx = \frac{\pi^{1/2} g^3 A_1}{6\eta A_2^{3/2}} \int_0^\tau p_b\,dt,$$

from which it follows that:

$$\int_0^\tau p_b\,dt = \frac{3\eta A_2^{3/2} l^2}{\pi^{1/2} g^3 A_1}.$$

The value of the integral must be less than that corresponding to the time for the wheel to be braked to the chosen slip criterion ($\geqslant 10\%$ in our example) under adverse conditions. This value depends on the moment of inertia of the wheel and is given by $K_b I \Delta\dot{\theta}$, where $\Delta\dot{\theta}$ is the permitted velocity change corresponding to the chosen slip ($> 10\%$) and K_b is a constant relating the braking torque to the braking pressure.

Provided the value of the integral satisfies the requirements, the fail-safe valve is cleared well before the anti-lock action could be called on to take place.

The anti-lock actuator described here, originally built for testing the electronic circuits of anti-lock systems, has worked well in field trials. Later versions have been constructed with flow control via edge orifices instead of annular channels to give a more viscosity-independent behaviour. Some of these actuators, incorporating many of the features described here, are now under further development elsewhere.

Summary. Description of an electro-hydraulic brake-pressure control, developed for testing electronic circuits for anti-lock braking systems for road vehicles. After an introduction to the problem and a short description of the contact between the tyre and the road surface on braking, a survey is given of the basic principles of anti-lock systems. Such a system consists of three main parts: a sensor for the angular velocity of the wheel, a processing unit, and an actuator that, if required, can reduce the brake pressure and increase it again (this has since been modified and improved). The actuator includes an electronically controlled solenoid valve, which reduces the brake pressure if wheel lock is imminent by causing brake fluid to flow out of the brake system, a pump that returns the fluid to the master cylinder, and a variable restrictor, which ensures that the rate at which the brake pressure increases again depends on the type of road surface. Various fail-safe arrangements are included to give trouble-free operation: if a leak occurs in the solenoid valve, or even if the valve stays open, normal braking is retained.

PHILIPS TECHNICAL REVIEW

VOLUME 36, 1976, No. 4

An experimental video-telephone system

E. A. Aagaard, P. M. van den Avoort and F. W. de Vrijer

'Mr. Watson, come here, I want you.' Those were the first words ever transmitted loudly and clearly by telephone. It was the 10th of March 1876 and Alexander Graham Bell was calling to his assistant for help, having splashed battery acid over his clothes while installing a new type of microphone. A hundred years have passed since then: the telephone is now commonplace and thought and interest have moved on — in the direction of a combination of speech and vision. The video telephone is still in an experimental stage; not because the necessary techniques are unknown but because of a desire to assess the value to the user of the various facilities it offers. A small experimental system has been in operation at Philips Research Laboratories for a number of years. Now the Netherlands Post Office and Philips have jointly set up a 'national' experimental network and linked it up with the original system. The article is mainly concerned with the system used at the Research Laboratories and gives a general description. A subsequent article will examine in greater detail the digital means used for synchronization and sound transmission.

Video telephone

The video telephone or videophone enables users not only to hear but also to see each other. The call gains considerably by this addition: users can see each other's facial expressions and show each other drawings, documents, etc. Contact is thus closer and conversations are possible which could not be conducted by ordinary telephone. The video telephone can even be used for conferences, so that participants do not have to travel.

The television picture accompanying the speech forms a valuable addition to the telephone call even if only the head and shoulders of the user are clearly imaged. Naturalness is still further enhanced if the normal handset is replaced by a loudspeaking telephone.

A video telephone of this kind can be achieved with existing techniques. For sufficient sharpness the picture has to have about three hundred lines; the video signal employed will then have a bandwidth of about 1 MHz. The transmission of signals with that bandwidth over existing telephone cables is possible over short distances. As in the existing telephone service, the connections for a video-telephone service have to be established through an exchange. Existing automatic telephone exchanges are not suitable for switching signals with such a large bandwidth; either they will have to be modified or special exchanges must be built.

In view of this situation it has been suggested that completely new broadband systems should be built, which could also be used for the transmission of data and facsimile and for the distribution of broadcast television. The existing telephone cables, however, could still be used for local video-telephone circuits.

To enable the technical problems associated with a video-telephone service to be studied and the utility of such a service to subscribers to be evaluated, a small experimental system of 20 stations was set up at Philips Research Laboratories in Eindhoven. Some of these stations are located in video-telephone kiosks for general use. The system, which went into operation in 1972, has now been incorporated in a larger experimental network with trunk circuits between The Hague, Leidschendam, Hilversum and Eindhoven (*fig. 1*), set up jointly by the Netherlands Post Office and Philips Telecommunication Industries [1]. This larger system, comprising a total of 65 stations, went into use in March 1974.

Ir E. A. Aagaard and Ir P. M. van den Avoort are with Philips Research Laboratories, Eindhoven; Dr F. W. de Vrijer is with Philips Research Laboratories, as a Scientific Adviser.

[1] L. J. W. van Loon, H. van der Hoff and S. J. A. Knijnenburg, An experimental video telephone network, Philips Telecomm. Rev. 32, 11-23, 1974.

This system is also intended for experimental purposes only and does not imply any decision concerning the future introduction of an actual service. It is quite possible that the video telephone will first go into normal service in government departments, commerce and industry, and not become available in the private sector until very much later.

It is difficult to predict the extent to which the need to travel will be affected by the existence of a reliable video-telephone service, but the question is an interesting one in view of the expected increase in personal travel. This is why the experiments include the holding of conferences by video telephone (*fig. 2*). At any rate users report that they find a video-telephone call comes somewhere between an actual conversation and a telephone call; perhaps there is some significance in the fact that people with a choice like using the video telephone for longer calls.

The account of the video telephone given here is a technical one. It deals with the subscriber's station and, briefly, with the method of sound transmission and synchronization. First, however, we shall discuss the signal standard chosen for the video telephone.

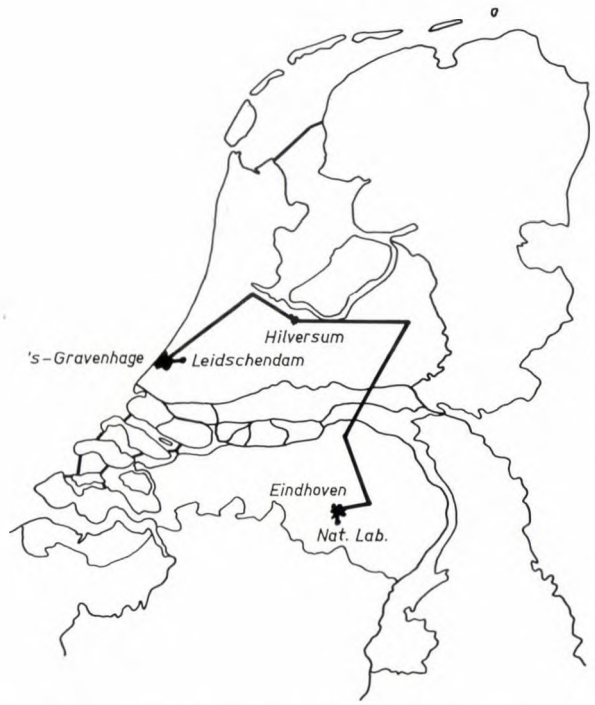


Fig. 1. The 'national' experimental video telephone network. Subscriber sets are installed at the Netherlands Post Office establishments at The Hague and Leidschendam and on Philips premises at Hilversum and Eindhoven, as well as Philips Research Laboratories (*Nat. Lab.*). The longer distances are spanned by radio links and coaxial cables. The total number of terminals is 65.

Choice of signal parameters

Picture

The bandwidth of a television signal is determined by the number of picture elements per second which it is desired to transmit. This number increases with the number of lines per picture and the number of pictures per second; for a given picture height it is also proportional to the picture width.

The bandwidth B is given by:

$$B = \tfrac{1}{2} n^2 \frac{1-p}{1-m} afK,$$

where n is the number of lines and a the ratio of the width to the height. The visible picture consists of fewer than n lines; a proportion pn occur in the field-flyback time. Initially, therefore, there are $n^2(1-p)^2a$ elements per picture. These have to be transmitted f times per second. The time available for this, however, is not continuous; for a fraction p of the field time and a fraction m of every line time (line-flyback time) no picture information is transmitted. This increases the bandwidth by a factor of $(1-p)^{-1}(1-m)^{-1}$. Finally, a correction has to be made for perceptual reasons. The line structure makes the resolving power



Fig. 2. Conference by video telephone. Conference stations have been installed at several points in the experimental network. In the experimental station shown here (at Philips Research Laboratories) only a single large monitor is used; the other conference stations employ several ordinary video-telephone sets.

of the eye in the vertical direction smaller. A subjectively equivalent reduction in the horizontal direction is also permitted; this is taken into account in the Kell factor K, which is smaller than unity. We have taken a value of 0.65 for K.

A picture rate of 25 per second is an obvious choice, since it is the same as that employed by TV broadcast stations. One advantage of this picture rate is that it is compatible with the use of fluorescent lamps, whose light emission has a ripple synchronous with the 50 Hz of the mains voltage. As in television, interlaced scanning, i.e. the scanning of first the odd lines and then the even lines, is used to avoid flicker. To ensure that a picture has a bandwidth of 4.8 MHz and can usually only be conveyed over a short distance on a telephone cable.

A picture consisting of about 300 lines ensures a good head-and-shoulders likeness of the speaker. The associated bandwidth is over 1 MHz. Such a picture can be compared with a quarter of an ordinary television picture (*fig. 3*). Unrepeatered transmission over existing local telephone cables is possible up to $1\frac{1}{2}$-2 kilometres; beyond that amplification will be necessary, especially for the highest frequencies. We opted for 313 lines because this number makes conversion to the broadcasting standard (625 lines) and vice versa relatively straightforward [2]. This decision was welcomed internationally for the same reason. A 625-line



Fig. 3. A picture of about 300 lines, as used for the video telephone, is equivalent to one-quarter of the picture used for broadcast television.

sideways movement does not make the speaker disappear immediately from the screen, the width of the picture is greater than its height; as in broadcast television the 'aspect ratio' is 4 : 3.

There is little freedom of choice for these various parameters, and it is consequently the number of lines which mainly determines the bandwidth. Here a compromise has to be struck between the picture quality and the transmission requirements. If existing telephone cables are used locally for the video telephone, two problems arise because of the high frequencies: attenuation, and crosstalk on other pairs. The permissible attenuation is limited by crosstalk, noise and other interference. To reduce these effects repeaters have to be introduced into the cable at regular intervals.

Sound

The limited bandwidth customary in telephony (300 to 3400 Hz) does not yield very satisfactory reproduction over a loudspeaking telephone. To improve the quality of the sound we extended the bandwidth to 5 kHz, so that the sound is now comparable for quality with that of medium-wave sound broadcasts. This

[2] M. C. W. van Buul and L. J. van de Polder, Standards conversion of a TV signal with 625 lines into a videophone signal with 313 lines, Philips Res. Repts. **28**, 377-390, 1973, and: Standards conversion of a videophone signal with 313 lines into a TV signal with 625 lines, Philips Res. Repts. **29**, 413-428, 1974.

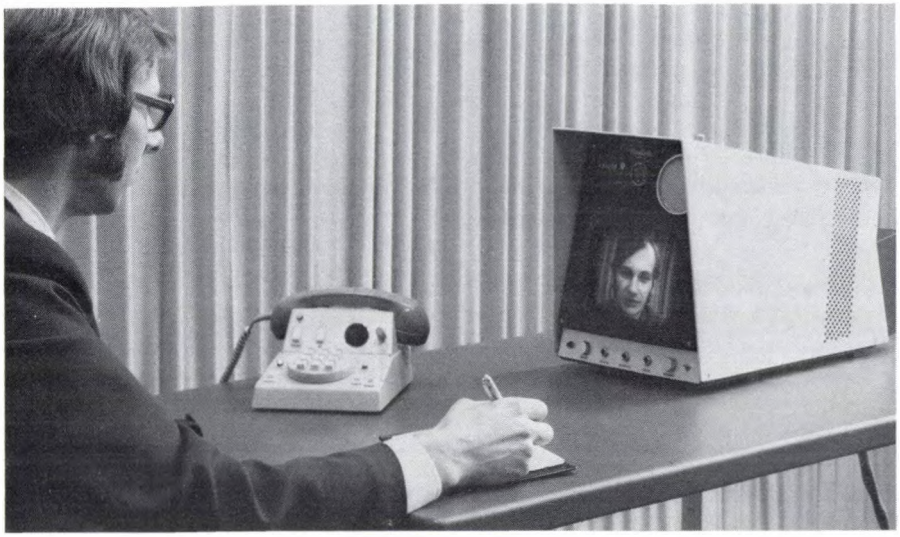


Fig. 4. Video-telephone set developed and in operation at Philips Research Laboratories. It consists of a picture unit, a control unit and a service unit. The picture unit incorporates both the camera (*top centre*) and the picture tube, as well as the loudspeaker. The control unit contains the microphone of the loudspeaking telephone system in addition to the pushbuttons for 'dialling' calls, and several switches for a variety of functions; if necessary, the loudspeaking telephone can be replaced by the handset. The service unit contains various supply circuits and also provides the matching to the telephone lines.

increase in bandwidth was possible because the sound is not transmitted along a telephone line in the normal way but is digitally encoded and then combined with the video signal.

The video-telephone set

The design of a video-telephone set requires a study of both the surroundings in which it is used (lighting, site, sound level, etc.) and the human factors that play a part in relaxed and natural conversation [3].

A video-telephone set for the experimental network was developed at the Research Laboratories (*fig. 4*); this was followed later by a design by Philips Telecommunication Industries (PTI) for the 'national' network (*fig. 5*). The sets differ in a number of minor technical features. They both consist of three units: the actual picture unit, the control unit and a service unit. The picture unit comprises the picture monitor, the camera and also the loudspeaker or loudspeakers. The control unit carries the pushbuttons for selecting the desired subscriber, a number of switches and the microphone of the loudspeaking telephone; it also incorporates a handset which can be used instead of the loudspeaking telephone. The service unit, which can be stored out of sight, houses the power supplies for the other two units and networks for matching to the telephone lines.

Location of the video-telephone set

When two people are seated at a table talking to each other their eyes are at approximately the same level and 120 cm to 2 metres apart. It is not practical to simulate this situation exactly with the video telephone; the picture unit would get in the way. In practice it will have to stand on the corner of a table or desk, and the viewing distance will then be 90 cm to 1 metre.

The picture size is related to this distance. It has been chosen such that the 300 or so picture lines are only just visible; larger dimensions would not help and could even be a nuisance. The picture height has to be one-sixth to one-tenth of the viewing distance [4], i.e. 9 to 17 cm. The picture height of existing 9-inch monitor tubes comes within this range and these have been used for the video telephone.

The picture is seen with minimum distortion if the picture screen is set at right angles to the direction of viewing, i.e. with a slight backward slope. This means, however, that annoying reflections from the ceiling lighting can occur in the slightly convex screen; the slope has therefore been decreased slightly (in the PTI version the entire unit can be tilted backwards a little).

Brightness and contrast

Unlike a television receiver the video telephone is used in very brightly illuminated surroundings. Bright lighting is in fact necessary for the camera. Without suitable precautions the brighter ambient light would lead to reflections in the picture screen, direct reflection on the smooth face of the picture tube and diffuse reflections on the phosphor layer. The images reflected from the picture screen are annoying and the phosphor reflections give a diffuse background light, reducing the contrast between dark and light parts of the picture.

To counteract the diffuse reflection from the phosphor layer the picture tube is made of dark-tinted glass which attenuates the light. The reflected light passes through the glass twice and the light from the picture only once so that the relative brightness of the picture is improved.

The direct reflections are suppressed by an optical filter system consisting of a polarizing filter with a birefringent layer, in front of the picture. The diffusely



Fig. 5. Video-telephone station set designed by Philips Telecommunication Industries (PTI) for use in the national experimental network. Like the set developed at Philips Research Laboratories (fig. 4) it consists of three units. The technical differences are only minor.

reflected light is also attenuated twice by a factor of 2, i.e. in total by a factor of 4, by the polarizing filter. Reflections from the filter itself are prevented by an anti-reflection coating.

The method used to suppress the direct reflections is illustrated in *fig. 6*. The polarizer *P* transmits only the incident light with a particular direction of polarization (vertical in the figure). In the birefringent layer *B* the optical axes are at angles of 45° to this direction of polarization. The polarized incident light can be considered as the sum of two equal components polarized along these axes. Light polarized along the *SS* axis is propagated less rapidly through the layer than the other component. The thickness of the layer is such that on leaving the layer the first component has a phase lag of a quarter-cycle in relation to the other component. The emergent light is thus right-circularly polarized.

After reflection from the front of the picture tube it is left-circularly polarized, becoming horizontal after its second passage through the birefringent layer. It therefore cannot pass through the vertical polarizer.

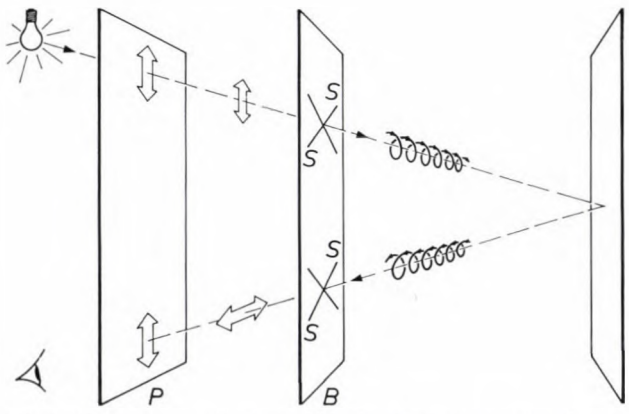


Fig. 6. Suppression of mirror (direct) reflections on the picture tube. A screen consisting of a polarization filter *P* and a birefringent layer *B* is placed in front of the picture tube. On passing through *P* and *B*, the incident light is circularly polarized. The direction of rotation of the polarized light is reversed on reflection at the picture tube. After its second passage through the birefringent layer the polarization is linear, but now in a direction perpendicular to that of the incident light; the polarizing filter therefore does not transmit the reflected light.

Eye contact

People engaged in an ordinary conversation look each other in the eye from time to time. With the video telephone the eyes are imaged slightly above the centre of the screen, which is where the lens of the camera should therefore be sited for eye contact. It has however been found that a deviation of the direction of viewing in the vertical plane goes unnoticed up to an angle of 3° and does not seem unusual until it reaches 12°. The camera lens is located just above the screen, in the middle. This is the direction in which any deviation is most readily tolerated; the deviation here is in fact 7.5°. If the lens were placed below the screen, the deviation would be less favourable, while the camera would also 'see' more of the brighter ceiling lighting.

Image plane, depth of focus

The camera lens ensures that an image plane situated about 1 metre in front of the camera is imaged sharply on the target of the miniature 'Plumbicon' camera tube used in the video telephone. The dimensions of the image on the target are about 8×6 mm. In the PTI stations the lens can be focused on objects at a greater distance to show a wall panel or blackboard.

The camera is equipped with a lens of continuously variable focal length (zoom lens). The picture area is largest at the smallest focal length. It must be sufficiently large to allow the user some freedom for small movements but it must not be unnecessarily large, either, because the face would then appear too small. The image plane can be shifted vertically by tilting the camera, over a range of $\pm 4°$. In the PTI station the camera is not tilted but the scanned part of the target

[3] R. R. Stokes, Human factors and appearance design considerations of the Mod II 'Picturephone' station set, Conf. Rec. 1968 IEEE Int. Conf. on Communications, Philadelphia, pp. 130-135.

[4] R. Theile, Fernsehtechnik, Springer, Berlin 1973.

in the camera tube is shifted, while zooming is effected electrically rather than optically, by reducing the scanned area of the target.

The depth of focus has to be sufficiently large to allow the speaker to move forwards and backwards slightly without the picture becoming blurred. However, any loss of definition less than that resulting from the line structure is imperceptible and therefore permissible. At the lowest focal length used, 16 mm, the depth of focus with the diaphragm fully open ($f/2.8$) is 70 cm, which is adequate. Zooming reduces the depth of focus considerably; see *fig. 7*. Fortunately, the zoom lens is used chiefly for displaying documents, when the reduced depth of focus is not a disadvantage.

The diaphragm is automatically controlled to ensure the correct illumination for all lighting levels. The adjustment is effected by comparing the output voltage from the camera tube with a reference voltage.

Graphic mode

Documents can be most conveniently shown to the person at the other end of the line if they can be placed flat on the table. A small retractable mirror with a lens attachment has been installed above the camera lens for this purpose. With this mirror in position the object plane coincides with the surface of the table. The maximum size of the object plane is then about 18×13.5 cm, and this can be reduced to a minimum of about 8×5.5 cm by zooming.

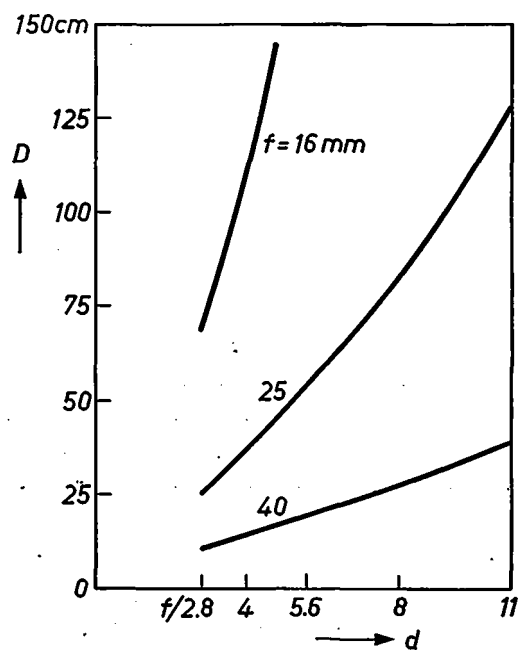


Fig. 7. The depth of focus D of the camera optics as a function of the relative aperture d of the lens. A the normal focal length (f = 16 mm) with the diaphragm fully open ($f/2.8$), the depth of focus is 70 cm, which is adequate for imaging the speaker. Zooming increases the focal length and reduces the depth of focus.

For good legibility at the far end of the connection documents have to be printed in a bold letter face. A printed page taken from this journal, for instance, would not be legible. Typed text is a borderline case. Drawings have to be simple and lines not too thin; a felt-tip pen gives good results.

Voice switch

Unlike the ordinary telephone system, the loudspeaking video-telephone system has separate speech circuits for the outgoing and incoming directions because the sound is combined with the picture to form a single signal. Without special precautions a closed circuit, partly electrical and partly acoustic, due to feedback from the loudspeaker to the microphone would be created for the sound and self-sustained acoustic oscillation ('singing') could occur. To prevent this, sound signals are admitted to only one direction of transmission at any given time. This is done by a voice switch in the video-telephone station, which compares the strength of the incoming signal with that of the outgoing signal and decides whether the outgoing signal will be transmitted. Switching occurs with a slight delay, chosen such that excessively frequent switching, e.g. due to background noises, is avoided, while loss of some speech elements due to the operating delay is not found to be annoying.

The voice switch derives its control signals from a compression-expansion (compander) circuit associated with the digital coding of the sound signals.

Screening

The combination of equipment needed to form the video-telephone station tends to give rise to certain characteristic faults. One of these is a disturbing whistle at the line frequency; another is the interference at the camera section from the deflection fields and voltages in the picture tube.

The line frequency used for the video telephone is 7825 Hz. Unless precautions are taken, a tone at this frequency is radiated as a result of magnetostriction in the transformers of the line-deflection and EHT circuits and in the deflection unit of the picture tube. This line frequency is in the middle of the audio band, unlike that which occurs in broadcast television (at 15 625 Hz) and which many viewers do not even notice. The main sources are the transformers; they are encapsulated in rubber in the video-telephone station and fitted with acoustic screening. This has the effect of making the EHT and deflection circuits somewhat bulky.

The magnetic and electrical crosstalk due to the deflection signals is another typical video-telephone problem, since the camera tube and receiver tube are alongside each other while the picture scanning is not

synchronous. The deflection fields in the picture tube have to be very strong to deflect electrons with an energy of 12 keV through an angle of 90° and some of this energy tends to find its way to the camera tube, where the electrons have an energy of only 300 eV. To prevent this, both the picture section and the camera section are fitted with magnetic screening.

Another problem is the crosstalk from the line-flyback pulses of the monitor to the camera preamplifier. Here again careful screening and the avoidance of earth loops were found to be necessary.

Control unit

Unlike the picture unit the control unit is within arm's reach. One of the switches on it can be used by the caller to display his own picture on the screen for monitoring purposes. The outgoing picture or sound can also be suppressed, and communication established with a secretary set. The zoom lens and, in the PTI station, the distance setting, are also controlled from this unit.

Signal transmission

Sound

As already mentioned, the sound signal is first digitally coded and then combined with the video signal. A four-wire circuit with separate outward and return paths for the sound, as well as for the picture, is therefore available, simplifying the incorporation of a loudspeaking telephone. Furthermore, because of the digital coding of the sound, the sound level and hence the operation of the voice switch are independent of the line attenuation and, unlike transmission in analog form via a separate line [5], the sound channel is relatively insensitive to disturbance by crosstalk. Another advantage of the joint transmission of the picture and sound is that no transmission-time differences can occur between them, as can happen when they have separate transmission paths.

The combination of the sound signal with the video signal is possible because of the 'gaps' in the video signal. No video information is transmitted during the line and field flybacks. There is therefore a time interval available at the end of each line (B in *fig. 8*) for the transmission of other information. Part B_1 is used for a signal which synchronizes the clock frequency at the receiving end with that at the transmitting end, and the remaining part B_2 is available for the sound signal. These signals are transmitted in digital binary form; the bit rate is 128 times the line frequency, i.e. 1.0016 Mbits/s. Time interval B contains 20 bits; the first 12 (a series '0101 . . . 01') are used for bit synchronization, the last eight for the sound. The sound signal is coded

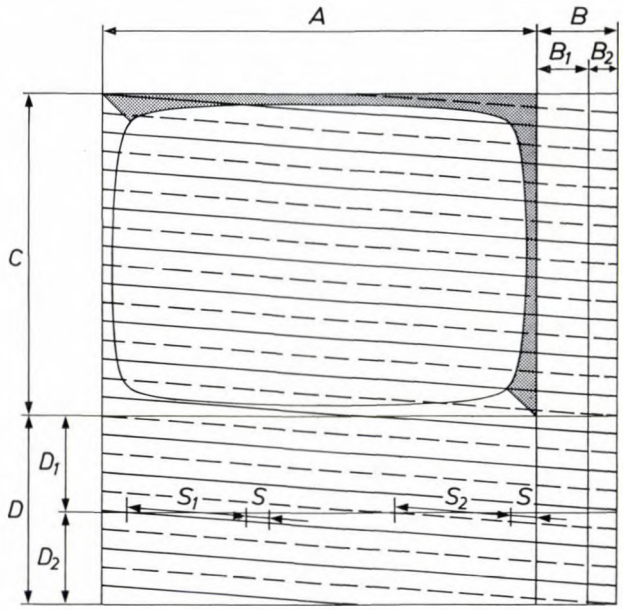


Fig. 8. Diagram illustrating how the picture is composed of lines (in reality more than are shown here). There are two fields; the odd field is written first (*dashed lines*) and then the even one (*solid lines*). A line period is $A + B$; video information can be transmitted during A, and B is used for bit synchronization (B_1) and digitized sound transmission (B_2). The lines contain the actual picture signal during period C and other information during the vertical flyback time D. In period D_1 a digital signal consisting of a continuous succession of 1's and 0's is transmitted; it is required at the start of a call in order to synchronize the clock signals and establish the black level. In period D_2 groups of 32 bits are transmitted for synchronization of the even field (S_1) and the odd field (S_2), together with dialling signals (S). The time still remaining in D_2 can be used to convey other information (e.g. call charges).

by means of delta modulation, employing signal companding with digital control [6]. With eight bits per line period there is a bit flow of 62.6 kbits/s. The bits are generated with fixed intervals and stored in a shift register for transmission in groups of eight at the increased bit rate of 1.0016 Mbits/s during the B_2 periods. The process is reversed in the receiver, resulting in a flow of equidistant bits.

Synchronization and signalling

The form chosen for the synchronization signals in the video-telephone network is different from that used in television broadcasting, where the amplitude of the sync pulses is about 40% of the distance between the black and white levels in the video signal; see *fig. 9*. This amplitude is inconveniently large for video telephony; the line repeaters can only amplify distortion-free up to a certain amplitude and as much of this signal space as possible must be available for the

[5] See the 'Picturephone' issue of Bell Syst. tech. J. **50**, 1971 (No. 2, pp. 219-709).

[6] J. A. Greefkes and K. Riemens, Code modulation with digitally controlled companding for speech transmission, Philips tech. Rev. **31**, 335-353, 1970.

actual video signal. Moreover, sync pulses of large amplitude have a marked tendency to cause interference on other circuits. For these and other reasons it was decided to employ a digital synchronization system with pulses of amplitude only 10% of the black-to-white distance; see *fig. 10*. These pulses are situated symmetrically about the black level, which is consequently easily identified. This also means that only 5% of the black-to-white distance need be reserved as signal space for the sync pulses.

All sync signals at the transmitting and receiving ends are derived from a high-frequency oscillator which also generates the bit-clock signals. The oscillator at the receiving end is synchronized with the transmitting oscillator by means of the pulses in period B_1. Signals S_1 and S_2 are used for field synchronization and can be easily accommodated in the vertical-flyback period. These signals are shown in fig. 8, where S_1 is the signal for even fields and S_2 that for odd fields. They are series of 32 bits; S_2 is the logical inversion of S_1. The pattern of this series is stored in the receiver and the flow of incoming bits is continuously compared with it. The bit series (in which the first 31 bits correspond to a period of a 'maximum-length sequence') is chosen in such a way that the chance that interference will cause the receiver to indicate a correspondence at the wrong time is at a minimum. This sync system is so insensitive to noise and other interference that synchronization is not lost until interference levels are reached at which the picture and sound have already become unacceptable.

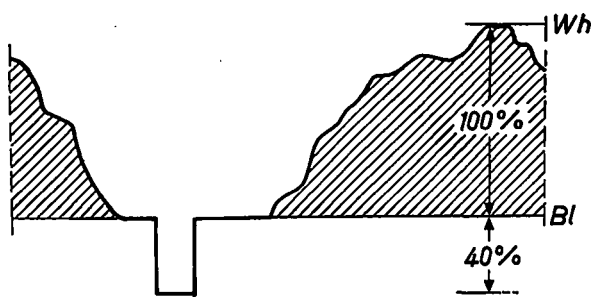


Fig. 9. Sync pulse in a broadcast-television signal. The amplitude of the pulse is about 40% of the distance between the white level *Wh* and the black level *Bl*. The pulse therefore occupies a considerable part of the signal space.

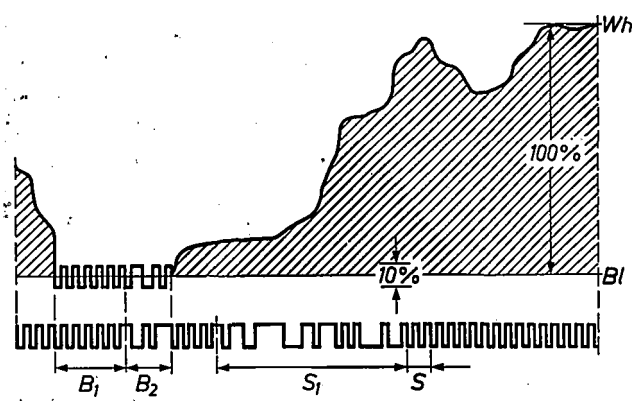


Fig. 10. Sync pulses in the video-telephone signal. Pulses for bit synchronization are transmitted in period B_1 (see also fig. 8) and the sound in digital form in period B_2. This is done both during the transmission of the picture (*above*) and during the vertical-flyback time (*below*). A line from the vertical-flyback time is shown in which the field-sync signals S_1 and the dialling information S occur. The peak-to-peak value of the sync signal is 10% of the distance between the white level *Wh* and the black level *Bl*; the additional demand on the signal space by the sync signal is thus only 5% of the black-to-white distance.

In the first part D_1 of the vertical flyback period D (fig. 8), a continuous succession of 0's and 1's is transmitted for the purpose of effecting clock synchronization at the start of a call and establishing the black level. Only after this synchronization can the signal S_1 and S_2 be decoded.

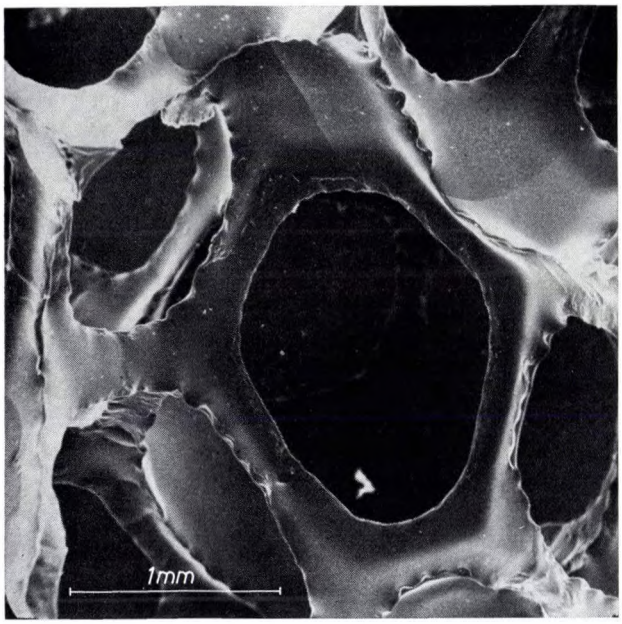
In the second part D_2 of the vertical-flyback period many bit places are still not used. Ten lines are suppressed per field; D_2 contains five of these in the experimental network but can be extended. Every line can accommodate 128 bits, 20 of which are needed for line synchronization and sound, which continue during the flyback period. That leaves 540 bit places per field, of which only 32 are used for field synchronization. A use can be found for the remaining bits. A modest start has been made in the experimental network by using 4 bits immediately after the 32-bit series to transmit the called subscriber's number, which is conveyed in digitally coded form to the exchange.

This still leaves an unused bit flow of approximately 25 kbits/s, but this could be employed to carry various kinds of information, not necessarily connected with the video-telephone call. An example of such information directly associated with the operation of the system would be a statement of the charge for a call or a warning that a third subscriber is calling.

Summary. Since 1972 a video-telephone service has been in use by way of an experiment at Philips Research Laboratories in Eindhoven; there are 20 terminals. Since March 1974 the service has been connected to Philips establishments at Eindhoven and Hilversum and the Netherlands Post Office at The Hague and Leidschendam. The system uses a 313-line television picture. At local level the video signal, which has a bandwidth of over 1 MHz, is conveyed on ordinary telephone lines, with repeaters inserted at intervals of 2 km. Radio links and coaxial cables are used for trunk traffic. A telephone exchange with special facilities is necessary. The sound signal, digitally coded, is transmitted during the line-flyback time of the television picture. A loudspeaking telephone is used. There are conference stations and video-telephone kiosks. The value placed upon a video telephone call by the users is found to be somewhere between those of an ordinary telephone call and an actual conversation. Video calls last longer on average than ordinary telephone calls.

Carbon foam

W. F. Knippenberg and B. Lersmacher



The article below deals with a new and promising variant of the familiar carbon and graphite materials. The special feature of this new material is that although its thermal and electrical conductivities are relatively low, owing to its low density, it can nevertheless be made very strong. Because of this combination of properties and its intrinsically inert chemical nature it shows promise for a wide range of applications from internal prostheses to heat shields for space technology.

The description of the new material is preceded by a general discussion of the structure and growth of graphite carbon. A series of photographs, showing the appearance of various forms of carbon, clearly indicates how the pattern of growth is affected by the conditions under which the growth takes place.

Introduction

Types of carbon and graphite, such as charcoal and the electrographites, have many technical uses — the carbons, for example, because of their adsorbent properties, and the graphites because of their high thermal and electrical conductivities and their resistance to corrosion, even at high temperatures. This makes the graphites a very suitable material for crucibles, arc-discharge electrodes, electrodes for high-temperature electrolysis, etc.

This range of well established types of carbon, which are still as important today as they ever were, has been widened in recent years by the introduction of some new types, in particular pyrolytic graphite and glassy carbon. Both materials and their potential applications have been the subject of previous articles in this journal [1].

In their structure these two new types of material represent two extreme forms that can be produced by carbon atoms with their tendency to form bonds in the layered lattice structure of graphite. Pyrolytic graphite has a density approaching that of single-crystal graphite (2.265 g/cm^3), and the crystallites from

Dr W. F. Knippenberg is with Philips Research Laboratories, Eindhoven; Dr B. Lersmacher is with Philips GmbH Forschungslaboratorium Aachen, Aachen, West Germany.

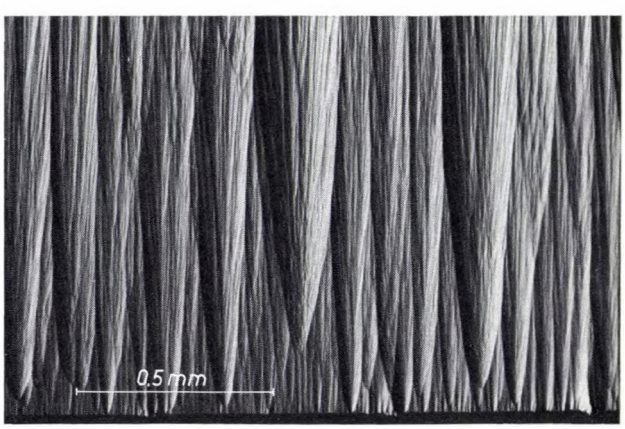
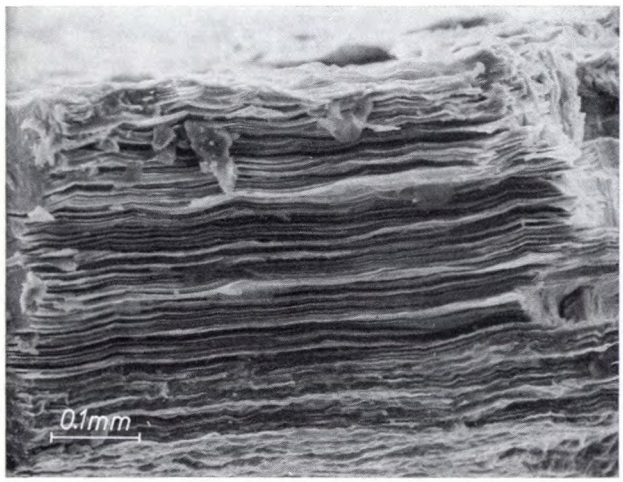
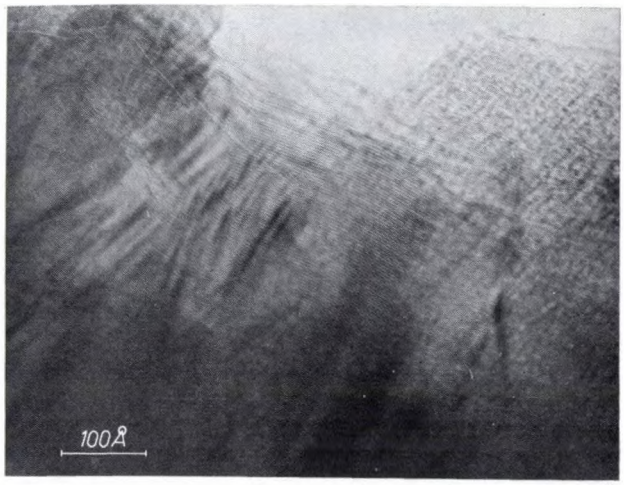


Fig. 1. Photomicrograph of a polished section of pyrolytic graphite, deposited at 2000 °C from propane on an electrographite substrate. The photograph was made with polarized light. The preferred orientation of the crystallites is apparent in the parallel paraboloidal cones. The cones are built up from curved graphite layers in which the line normal to these layers (the *c*-axis) and the direction of the cone axis differ only very slightly. The relief in the photograph arises because the sections of the cones, which in reality lie in a single plane, reflect the polarized light with different intensity, since the cone intersections represent different cross-sections of the crystal lattice.

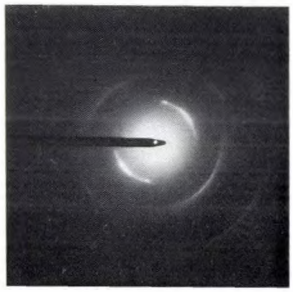
[1] Pyrolytic graphite is discussed in: W. F. Knippenberg, B. Lersmacher, H. Lydtin and A. W. Moore, Philips tech. Rev. **28**, 231, 1967, and glassy carbon in: Philips tech. Rev. **31**, 369, 1970, and also in: B. Lersmacher, H. Lydtin and W. F. Knippenberg, Chemie-Ing.-Technik **42**, 659, 1970.



a

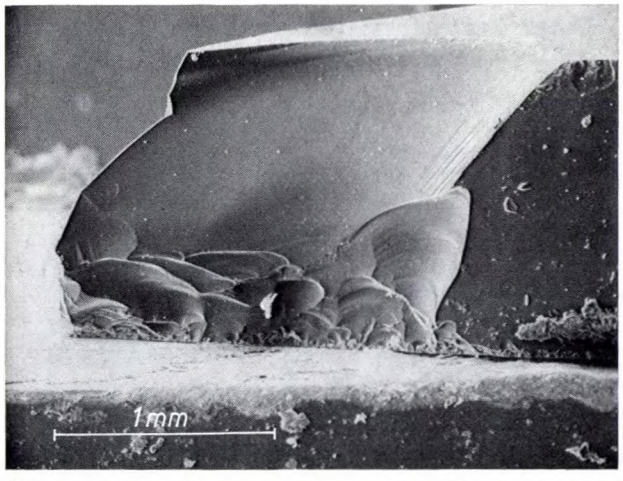


b

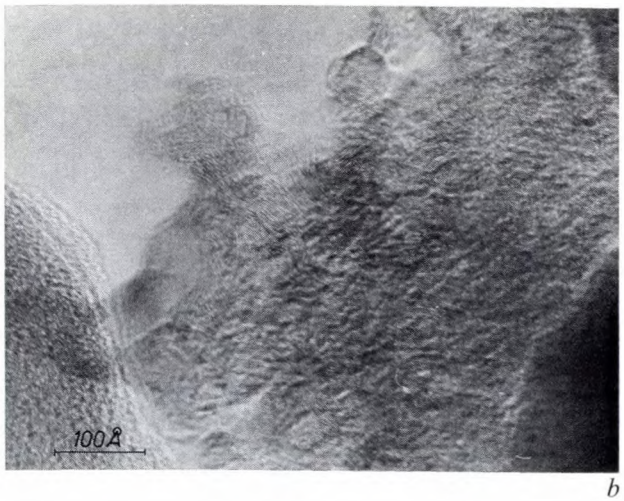


c

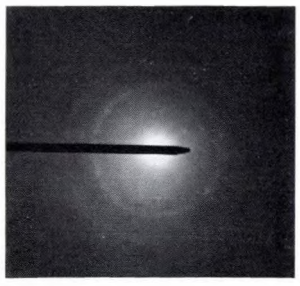
Fig. 2. *a*) Scanning electron photomicrograph of a fracture plane in pyrolytic graphite. The fracture lines clearly show the layer structure of the material. *b*) Transmission electron photomicrograph of pyrolytic graphite. The ordered atomic carbon layers can be seen that give rise to a sharp electron diffraction pattern (*c*).



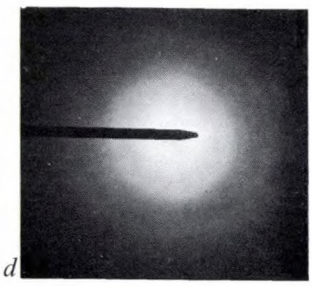
a



b



c



d

Fig. 3. *a*) Scanning electron photomicrograph of a fracture surface of glassy carbon made by pyrolysis of phenol-formaldehyde resin on heating to 1400 °C. The fracture surface shown is typical of glassy material ('conchoidal' fracture). *b*) Transmission electron photomicrograph of a splinter of this glassy carbon. In parts of the photograph the atomic carbon layers are visible, forming entangled skeins, whereas in other parts there is hardly any structure. The structured areas of the material correspond to an electron diffraction diagram as in (*c*), the other areas to a diffraction pattern as in (*d*).

which it is built up exhibit a strong preferential orientation — compared with the crystallites in electrographites — giving this material certain properties that closely resemble those of the ideal graphite crystal (*figs. 1* and *2*). Glassy carbon has a much lower density (1.5 g/cm^3), owing to its much less ordered ('paracrystalline') structure. Even though such order as is present extends only over a short range, strong carbon-to-carbon bonds give strong cohesion to the 'crystallites' (*fig. 3*) — a situation different from that in the conventional carbons. The lower density is associated with relatively low thermal and electrical conductivities.

Both materials typically possess mechanical properties superior to those of conventional types of carbon and graphite, i.e. higher values of compressive strength, tensile strength and stiffness. A practical limitation in the use of these materials, especially glassy carbon, is that they cannot be made very thick.

This article deals with a further development, which has given material that may be described as 'carbon in macroporous form', or briefly as 'carbon foam'. It is a very light material, made by pyrolysis — thermal decomposition — of certain foamed polymers. Unless special precautions are adopted in the pyrolysis, the carbon foam obtained is generally of low mechanical strength, and until now has only been of use for thermal insulation. We have developed a carbon foam, which we call 'glassy carbon foam', that does have relatively high mechanical strength (*fig. 4*). This new variant continues the development, initiated with glassy carbon, towards lighter and stronger types of carbon, thus opening up further prospects of completely new applications.

Glassy carbon foam is prepared by impregnating a polymer foam of reticular structure (*fig. 5*) with a carbon-rich liquid, such as a solution of a phenol-formaldehyde resin. Heat treatments are then performed to expel the solvent and carbonize the impregnated porous material; under certain conditions a material emerges that combines very high porosity with great strength, because the introduced resin phase has acquired the character of glassy carbon during the pyrolysis.

In attempting to produce a material that is both light and strong a compromise always has to be found in a case like this. Adding more glassy carbon will make the material stronger but at the same time heavier. As we shall see presently, the density determines many other properties in addition to the strength. The density depends not only on the porosity of the original material but also on the degree of impregnation and on the method of pyrolysis. The range of variation is between 0.1 g/cm^3 and 1.0 g/cm^3 (compared with 1.5 g/cm^3 for solid glassy carbon). This corresponds to a variation in compressive strength of about 50 to about 25 000 N/cm^2. Although these compressive strengths are greater than those of 'non-reinforced' carbon foam (whose compressive strength is at the most 20 to 30 N/cm^2), they are of course less than those of glassy carbon, which range from about 26 000 to 50 000 N/cm^2. Set against this lower compressive strength is the advantage that glassy carbon foam can be made at any thickness, whereas for compact glassy carbon the upper limit is in practice about 3 mm.

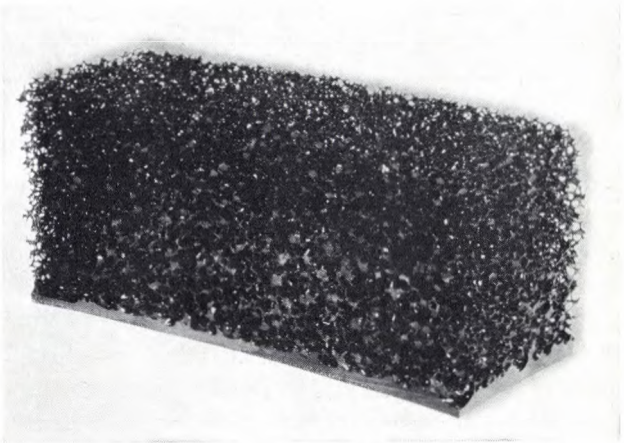


Fig. 4. A piece of glassy carbon foam (12 cm long).

The properties of a carbon material depend to a very great extent on its structure — taking the word in its

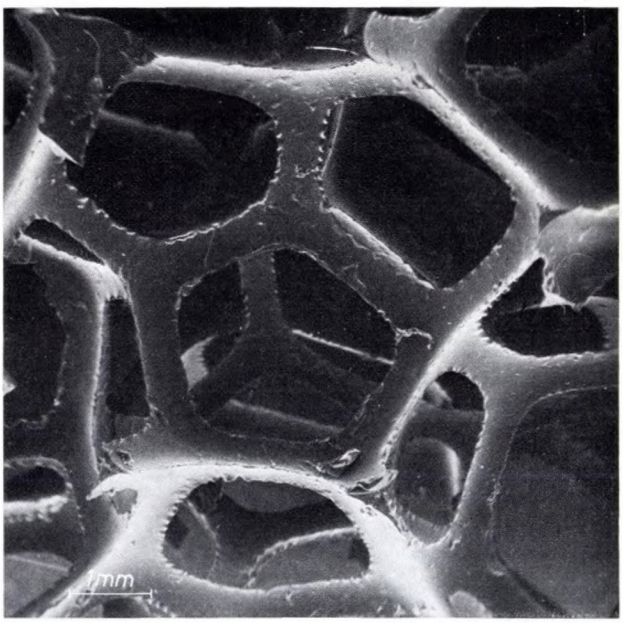


Fig. 5. Scanning electron photomicrograph of foamed polyurethane used as the starting material in the preparation of glassy carbon foam. Comparison of this photograph with the end product illustrated in the title photograph shows that the open reticular structure originally present has partly closed up in the end product: cell walls are now visible as well as cell filaments.

widest sense — which in turn is closely related to the manner in which the material is produced, and also of course to the ways in which the carbon can grow under the given conditions.

In the next section we shall deal first with the structure and growth of carbon in general, before going on to discuss the preparation, properties and applications of glassy carbon foam, the real subject of this article, in the later sections.

In view of the accessibility and size of the internal surface of this material, special attention will also be given to the possibilities of converting it into a metal-carbon composite or a carbide, completely or partly, while retaining its porosity.

Structure and growth of carbon

Apart from diamond modifications, produced at a pressure between 10^4 and 10^5 bars, graphite is the stable solid modification of the element carbon. The characteristic feature of graphite is its hexagonal layer structure, illustrated in *fig. 6*: planar layers of carbon atoms are stacked in a regular configuration of hexagons in which the smallest atomic spacing is 1.42 Å (compared with 1.54 Å in the tetrahedral configuration of diamond). Half the atoms of each layer have the same position as those in the next layer above: the other half lie beneath the centres of the hexagons formed by the atoms of the next layer. This gives two equivalent stacking alternatives. In the stable graphite modification the atoms of the third layer have the same positions as those in the first layer. The other alternative, in which the atoms of the fourth layer have the same positions as those of the first, is the metastable rhombohedral modification of graphite.

In graphite the distance between two adjacent layers is 3.35 Å. The small atomic spacing — and hence the strong bonding — *inside* a layer, and the weak bonding *between* the layers have long explained the ease with which the layers of graphite can be displaced relatively to one another and the graphite crystal cleaved parallel to these layers (see also fig. 2*a*).

It is known from X-ray and electron-diffraction patterns that both periodic and non-periodic stacking variants frequently occur in graphite. There are also variants in which the layers are rotated in relation to each other (*fig. 7*). The frequent occurrence of such stacking variants, and also of dislocations and twin-

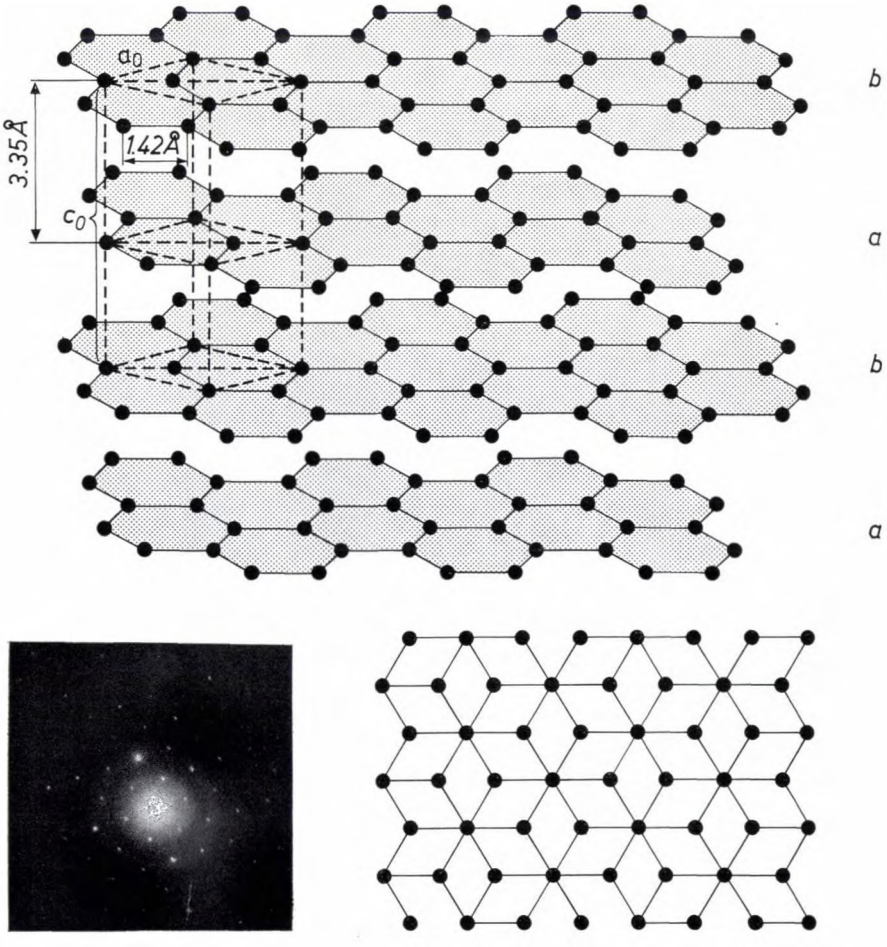
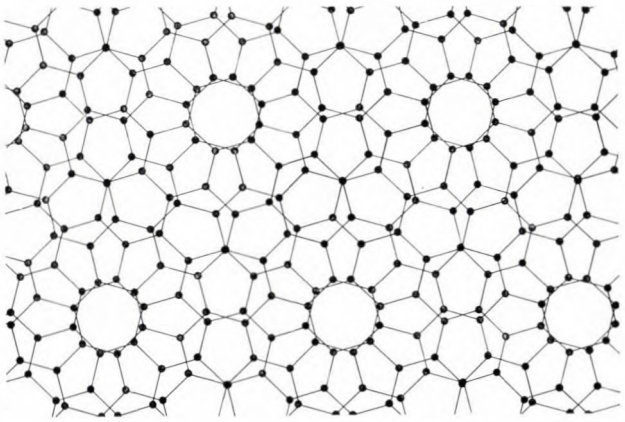
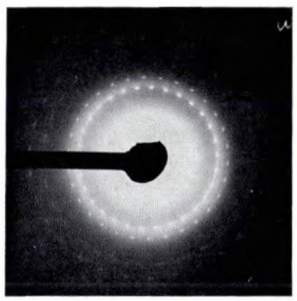


Fig. 6. *Above:* schematic representation of the atomic carbon layer structure of the stable hexagonal modification of graphite. For the projection perpendicular to the layers the structure has the appearance shown below. *Bottom left:* electron-diffraction diagram corresponding to this structure.



a



b

Fig. 7. *a*) One of the possible deviations fom the ideal structure: a rotation of the atomic carbon layers relative to one another through an angle of 30°. The diagram illustrates two layers rotated in this way, in a projection perpendicular to the layers, clearly demonstrating the way in which periodic density centres occur [2]. *b*) Electron-diffraction diagram corresponding to structures of this type.

ning, all indicate that the energy differences associated with these deviations from the ideal structure are small [3].

The wide latitude for variation in the lattice configuration, as outlined above, accounts for the ease with which carbon deposits based on the graphite lattice are able to adapt, in their nucleation and growth, to a compulsion imposed upon them to assume a particular macroscopic form.

The growth of carbon starts with nucleation, possibly on a substrate, and it may be fed from a gaseous, a liquid or a solid phase, called the mother phase, in which carbon is present in a certain degree of 'supersaturation'.

The most unperturbed growth takes place when there is homogeneous nucleation in an isotropic mother phase. In the 'ideal' case where the building units of the lattice being formed are able to take up their equilibrium positions completely freely, a morphologically ideal single crystal emerges. In practice, ideal growth does not take place, and if for example the transport in the mother phase is the slowest step in the growth process, and if furthermore the 'building elements' have little possibility of migrating on the nucleus, the pattern of the material and heat flows will influence the manner of growth. It will do so all the more strongly since, as we have seen above, carbon deposits are highly flexible and adapt easily to an external constraint. The manner of growth is also influenced by the substrate in the case of heterogeneous nucleation. Here not only the nature and form of the substrate are important but also the fact that the nuclei on the substrate are fixed in relation to one another and therefore interact. The same applies when the growth takes place, without a substrate, in a solid mother phase.

The influence of the growth conditions is reflected not only in the form of growth but also in the degree of perfection of the three-dimensional atomic lattices formed, and in the way in which these differently nucleated structures are interconnected.

Some of the various modes of carbon growth are depicted in *figs. 8* to *11*.

The modes of carbon growth can most easily be studied on dispersed forms ofc arbon, such as carbon precipitates in gases and in metals. It is much less easy to study the growth process of a compact phase like the one that occurs in the formation of glassy carbon foam. Most success has so far been achieved in investigations on simple *inorganic* substances such as silicon carbide. In the pyrolysis of this material the SiC crystals originally present are found to change to porous graphite crystals while preserving their outward shape [4]. Studies of the much more complex pyrolytic conversions in *organic* substances have previously resulted only in some broad understanding of the chemistry of the conversions and of the structural changes taking place during pyrolysis [5], but the details of the growth mechanism are still uncertain.

Preparation of glassy carbon foam

The preparation of carbon foam has until now generally been based on thermosetting, not thermoplastic, materials. With rising temperature thermosetting materials become hard because of the occurrence or further expansion of a three-dimensional — and irreversibly cross-linked — network of polymers. Thermoplastic materials, on the other hand, soften as the temperature is increased; they usually consist of linear polymers that do not become cross-linked under the influence of heat. Phenol, urea and epoxy-resin foams all belong to the first group, whereas polyvinyl-chloride, polyester, polyurethane and acrylonitrile-butadiene-styrene foams are included in the second group.

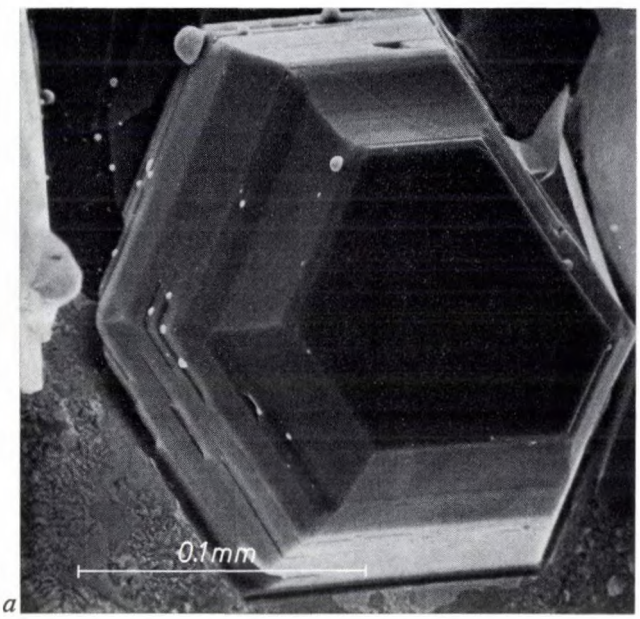
Whereas the thermosetting materials can often be carbonized as desired without altering their shape, the

[2] See for example D. D. Double and A. Hellawell, Acta Metallurgica **17**, 1071, 1969, and W. Bollmann, Crystal defects and crystalline interfaces, Springer, Berlin 1970.

[3] S. Amelinckx, P. Delavignette and M. Heerschap, in: P. L. Walker, Jr. (ed.), Chemistry and physics of carbon **1**, 1, Dekker, New York 1965.

[4] A. J. van Bommel, J. E. Crombeen and A. van Tooren, Surface Sci. **48**, 463, 1975.

[5] See for example E. Fitzer, K. Müller and W. Schäfer, in: P. L. Walker, Jr. (ed.), Chemistry and physics of carbon **7**, 237, Dekker, New York 1971, and J. L. White, G. L. Guthrie and J. O. Gardner, Carbon **5**, 517, 1967.



a

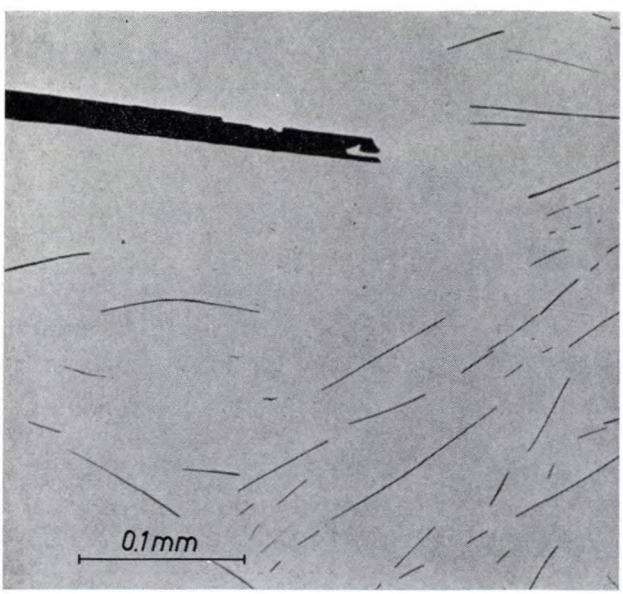


b

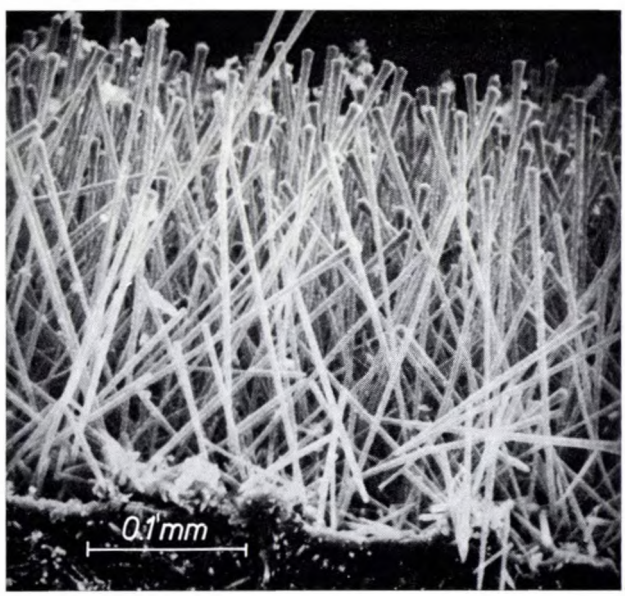
Fig. 8. There are various modes in which a carbon-containing system may yield to the growth compulsion arising from carbon supersaturation. The 'ideal' case is that in which the deposited carbon atoms are completely free to take up their equilibrium positions. The form of graphite crystal occurring under these conditions, as can be predicted from estimated surface energies or numbers of dangling bonds at the surface, is that of the *tabular* crystal (*a*). This has the general shape of a hexagonal prism, with the major basal plane parallel to the atomic carbon layers, bounded by first-order prismatic and pyramidal planes. However favourably the growth conditions may be chosen in practice (low supersaturation, high temperature, etc.), these forms of graphite are nevertheless seldom encountered, and never in dimensions of more than a few millimetres.

The fine example shown here was obtained from an iron melt prepared by Ir L. Kalvelage and Dr J. Pötschke at the Institut für Metallurgie der Kernbrennstoffe und Theoretische Hüttenkunde of the Technische Hochschule Aachen.

A much more commonly encountered growth mode, which is related to the first one, is *lamellar* growth (*b*). The plane of the lamellae is again parallel to the atomic carbon layers, but the ratio of diameter to thickness may be, for example, 10^4 times greater than that in the tabular crystal. In the case shown here it is clear that the growth kinetics, which particularly favour rapid growth in the direction of the strong carbon bonds, completely overshadow the thermodynamic factors. These lamellae also orginated from an iron melt, where different growth conditions evidently prevailed.



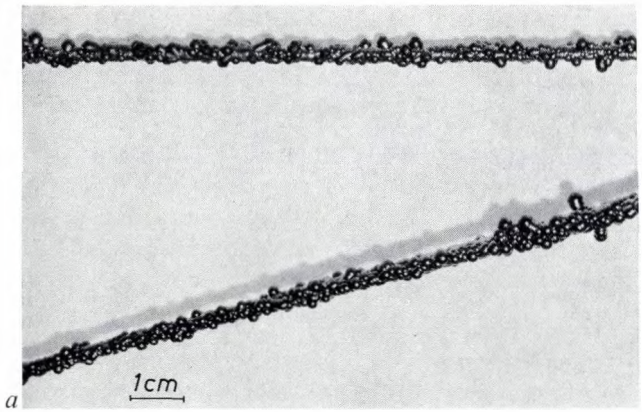
a



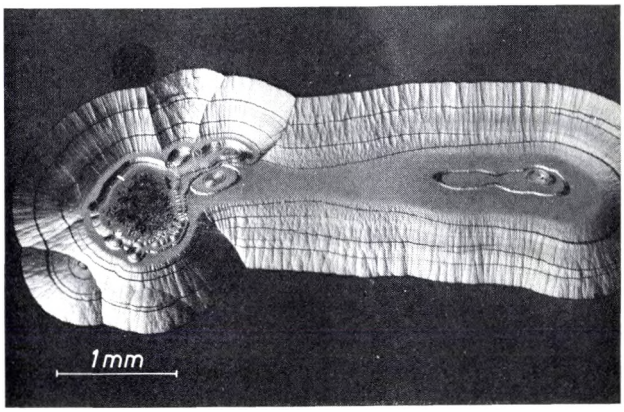
b

Fig. 9. Growth-kinetic factors even more dominant than those shown in fig. 8*b* are manifested in this picture in the form of whiskers, either as 'ribbons', which expand in one dimension in the direction of strong C-C bonds in the atomic carbon layers (*a*), or as 'columns', which grow in the direction of the weak bonds between the layers (*b*). As in all whisker growth, both variants must have been caused either by accelerated growth at the top or by inhibition or blockage at the side walls. In the first case the growth, once started at the top, keeps gaining on the growth elsewhere, even in crystallographically equivalent directions. This may be the result of local catalysis or of the greater concentration always present near the top. In the second case the inhibition or blockage is due to adsorption of impurities, which takes place more slowly at the top than the growth at that location. The situation shown in (*a*) is an example of the first case, i.e. accelerated growth at the top upon the growth of ribbons from an iron melt, while (*b*) is due to inhibition at the side walls caused by adsorption of CO, when columns are grown from CO gas at 2000 °C [6].

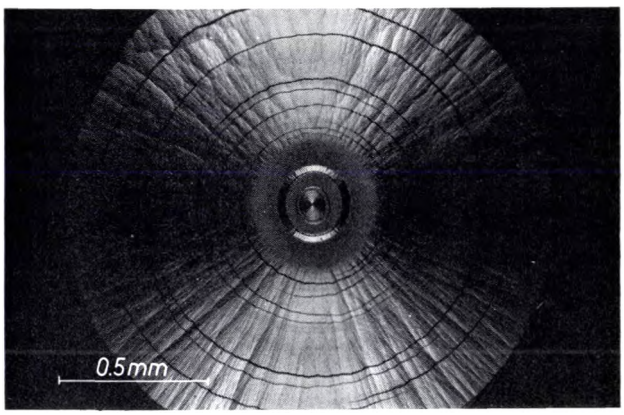
[6] See for example H. B. Haanstra, W. F. Knippenberg and G. Verspui, J. Crystal Growth **16**, 71, 1972.



a

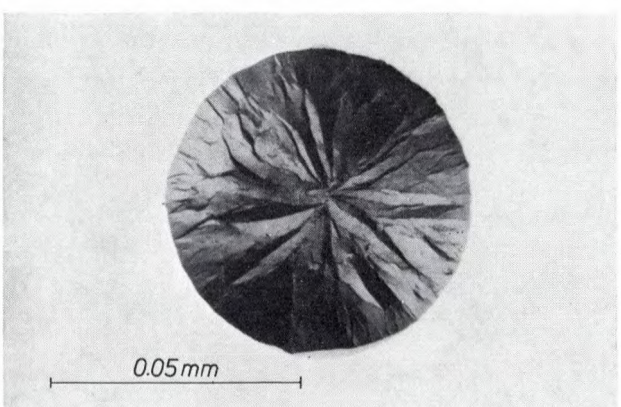


b

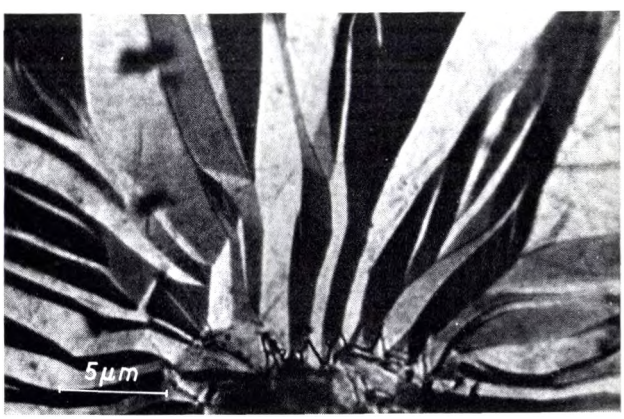


c

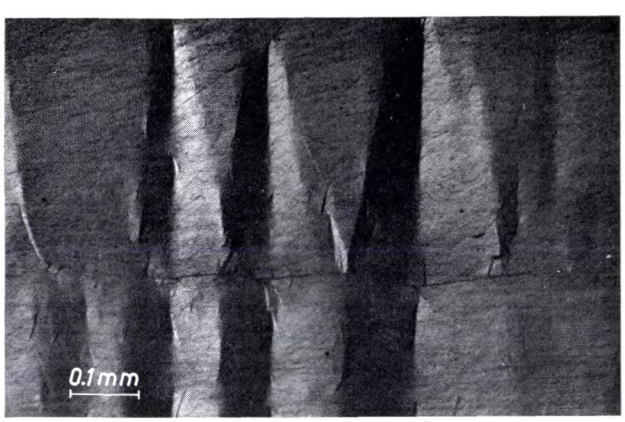
Fig. 10. Owing to the ease with which defects can be introduced, and effects such as twinning can arise, carbon layers faithfully follow all the irregularities in the substrate surface. In (*a*) it can be seen how two carbon threads (diameter about 1000 μm), consisting of 10 000 intertwined fine filaments, are completely enveloped by a layer of pyrolytic graphite, produced from the gas phase at 2000 °C. The photograph also shows how neatly the tops of projecting unravelled filaments (diameter about 7 μm) have been rounded off. In (*b*) and (*c*), which are longitudinal and transverse sections through a single filament respectively, it can be seen from the fracture lines that the atomic carbon layers in this process are always parallel to the substrate surface. These photomicrographs, like that of fig. 1, were taken in polarized light, so that the characteristic growth cones are again visible.



a



b



c

Fig. 11. Easy twinning when there is a strong growth compulsion and little opportunity for the 'building elements' to migrate also creates the possibility of a quasi-isotropic mode of growth, generally known as *nodular* growth. This can give rise to spherical nuclei which, given the appropriate growth conditions, could grow out in the manner illustrated in fig. 10, thus resulting in the formation of 'globular' graphite in a metal melt. The photomicrograph in (*a*) was taken in polarized light and shows a cross-section of a globular graphite sphere grown in a palladium melt. Different though the outward form is from that illustrated in fig. 8*b*, the underlying mode of growth is again essentially lamellar, although here, because of the spherical nucleation, it gives concentric shells. Although in this respect the result resembles the carbon deposition around a thread as shown in fig. 10, the difference between this pyrolytic graphite from the gas phase and globular graphite from a metal melt is also particularly striking: in the metal-melt case, conceivably owing to the pressure in the melt, a degree of orientation is obtained (see detail magnification (*b*), received from Dr B. Knook, Kamerlingh Onnes Laboratory of Leiden University), which has previously only been obtained after hot pressing of pyrolytic graphite (see *c*). The photomicrograph in *c* was taken in polarized light and shows a polished section of hot-pressed pyrolytic graphite, treated at temperatures above 3000 °C and at pressures of 100-500 bars. It can be seen that, as a result of this treatment, the variation in orientation originally present in each growth cone (see fig. 1) has largely disappeared.

thermal treatment of thermoplastics usually requires special precautions to ensure that the polymer does not depolymerize and evaporate, or melt, before becoming carbonized in the required manner.

The process that we use to form the carbon skeleton for the glassy carbon foam is nevertheless based on foamed thermoplastic, in particular on foamed polyurethane. The reason for this is that advantage can then be taken of the exceptionally wide range of possible variations in the cellular (or reticular) structure of these particular foamed materials [7].

The foamed polyurethanes have pores with an average diameter of 0.1 to 5 mm, depending on the degree of hardness obtained in the preparation. The 'softer' kinds exhibit the desired structure consisting of 'filamentary cells', which provide an open skeleton whose pores are readily accessible, and to which it is therefore easy to apply a 'second phase'.

Before the foamed material is impregnated with the carbon-rich liquid that will supply the second phase on pyrolysis, it must first be stabilized to ensure that the impregnating agent does not cause the material to swell, soften and eventually collapse.

Preliminary treatment: stabilization of foamed thermoplastics

Stabilization is sometimes effected by using gas oxidation to cross-link the starting polymers. We adopted a faster method by applying a protective layer to the thermoplastic filament cells by means of a liquid treatment. In this method of 'stabilization' an open foamed polyurethane is immersed at 80 °C in a solution of polyvinyl alcohol (5%) in water, and the water is then evaporated at 100 to 130 °C. The effect of this is to envelop the original cell filaments with a film of polyvinyl alcohol, which does not readily dissolve in the impregnating agent used in the next preparation step for the formation of glassy carbon.

Impregnation: introduction of thermosetting material

Since the purpose of the impregnation is to coat the material with a film of carbon-rich material capable of being converted into glassy carbon by pyrolysis, the coating substance will have to be a suitable type of thermosetting material [8].

The impregnation was usually carried out with phenol-resin solutions in ethyl alcohol. The amount of resin to be used can be regulated by the concentration of the impregnating agent and the number of times the impregnation is repeated. The more concentrated — and viscous — the impregnating agent, the less often the impregnation need be repeated. We have found that it is preferable to repeat the impregnation many times with an agent of low concentration and hence low viscosity, because this is the only way to achieve sufficient homogeneity and preservation of the open structure.

Before each fresh impregnation the layer previously applied has to be consolidated in a drying and polymerization process, for example by keeping everything for 20 to 30 minutes at 130 °C.

Pyrolysis and purification

A special method of pyrolysis is also required to convert the applied resin in the glassy carbon [8]. This is because the glassy carbon possesses the feature, unique among all the carbon products prepared from organic substances by pyrolysis, that it owes its great strength to the fact that the 'graphite crystallites' are formed during pyrolysis *in situ*, while retaining their mutual cohesion through strong C-C bonds (see also fig. 3*b*). This implies that the pyrolysis must be carried out in such a way that the decomposition products do not escape abruptly.

After undergoing the preliminary treatment described above, the material is subjected to a heat treatment in two stages [9]. In the first stage, which includes the carbonization, the temperature is gradually raised in an inert atmosphere to 900-1000 °C (see *fig. 12*, curve *I*). The material is now completely carbonized, but it still contains impurities, such as auxiliary materials used in the polymer preparation and hydrogen, one of the principal decomposition products of the pyrolysis process.

The second heating stage, designed to remove the impurities, is carried out in vacuum, and the temperature can now quickly be raised to 1600 °C or higher (curve *II* in fig. 12).

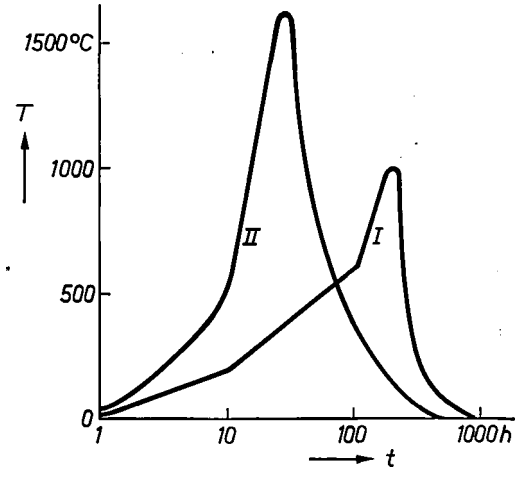


Fig. 12. Heat treatment for the pyrolysis of a resin-coated foamed polymer to produce glassy carbon foam. After an initial stage, in which the temperature varies as a function of time as illustrated in curve *I*, the material is subjected to a second treatment as shown by curve *II*. The first step causes carbonization, the second step is designed to remove impurities.

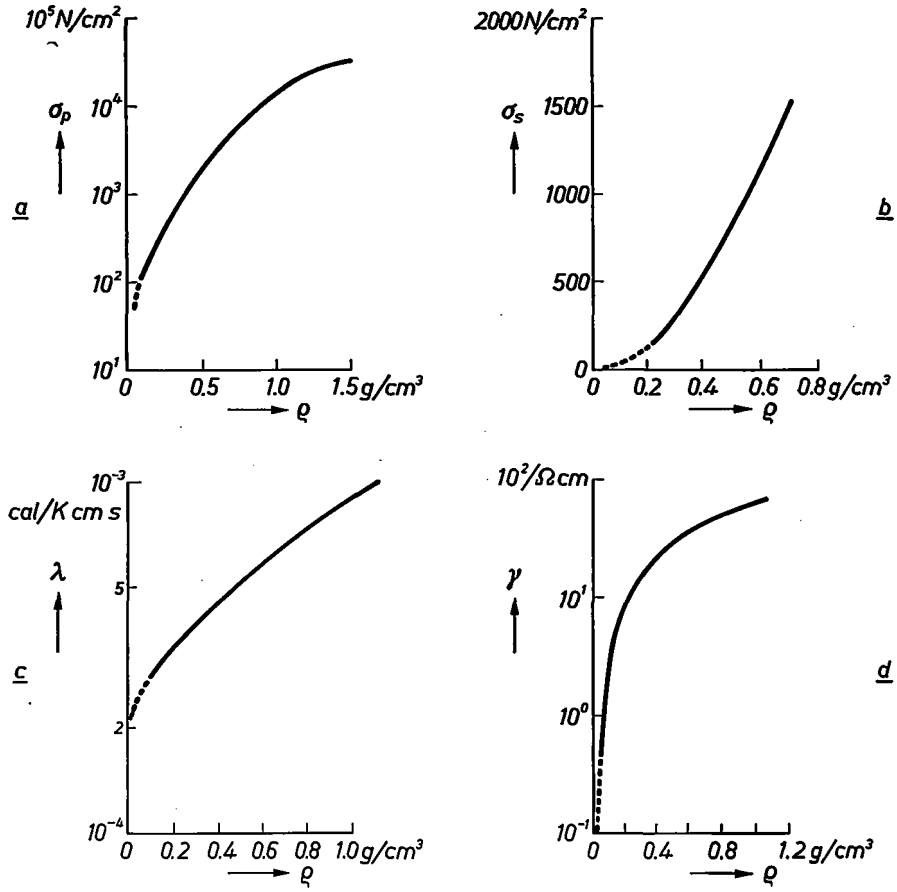


Fig. 13. The graphs in (*a*), (*b*), (*c*) and (*d*) present the compressive strength σ_p, the shearing strength σ_s, the thermal conductivity λ and the electrical conductivity γ respectively as a function of the density ϱ. The dashed lines relate to values for non-reinforced carbon foam.

The very gradual nature of the first heating stage is necessary to give the decomposition products an opportunity to escape by slow diffusion processes. This ensures that the structure of the material is not destroyed, giving a fragmented product.

The sluggishness of this diffusion sets the limit to the thicknesses in which glassy carbon can be produced. The porous structure of carbon foam however, reduces the severity of the problem: the glassy carbon applied is still thin, but the product as a whole is not.

Properties of glassy carbon foam

The properties of glassy carbon foam are directly connected with the density. The variation of compressive and shearing strength with density is shown in *fig. 13a* and *b*. It should be noted here that these properties are also affected by the pore structure (size, shape and concentration), so that the values given should only be regarded as averages. A suitable compromise between strength and lightness can be found in the density range from 0.75 to 0.3 g/cm³, which corresponds to a pore volume of 50 to 80%. The thermal and electrical conductivities as a function of density are shown in fig. 13*c* and 13*d*.

The dashed line in the figure indicates the values found for non-reinforced carbon foam. It may be concluded from this that, if the density of glassy carbon foam is increased by a factor of 10 over that of non-reinforced carbon foam, the strength can be increased by no less than a factor of 100. This 'asymmetry' partly accounts for the gain in strength offered by our glassy carbon foam.

Another improvement offered by glassy carbon foam as compared with ordinary carbon foam is found in its resistance to oxidation (*fig. 14*). Although oxidation of both glassy carbon foam and non-reinforced carbon foam starts at about the same temperature, irrespective of the density, it proceeds more slowly with rising tem-

[7] See for example C. J. Benning, Plastic foams, volumes I and II, Wiley, New York 1969.

[8] Since present knowledge of the pyrolytic growth process is still scanty and superficial, the choice of the basic chemical materials and the form of pyrolytic treatment was based mainly on experiment. Variation of these factors leads to very great changes in the properties of the applied layer of glassy carbon.

[9] We shall not consider the technical details of the carbonization process here. This is conventional, as described in: A. K. Thoeni, G. K. Baker and R. P. Chartoff, Bendix tech. J., Spring 1969; B. Lersmacher, Hochtemperaturfester Schaumstoff, Vortragsmanuskript DGLR-Tagung, Bremen 1969; American patents Nos. 3387940, 3121050 and 3342555; French patent No. 1388818.

perature in the case of the denser material. A contributory factor here is the slower gas diffusion with smaller pores. It should also be noted here that this resistance to oxidation can be substantially improved by applying a thin layer of pyrolytic graphite to the glassy carbon. A product obtained in this way, which could be called foamed graphite at the correct ratio of pyrolytic graphite to glassy carbon, will not be dealt with here.

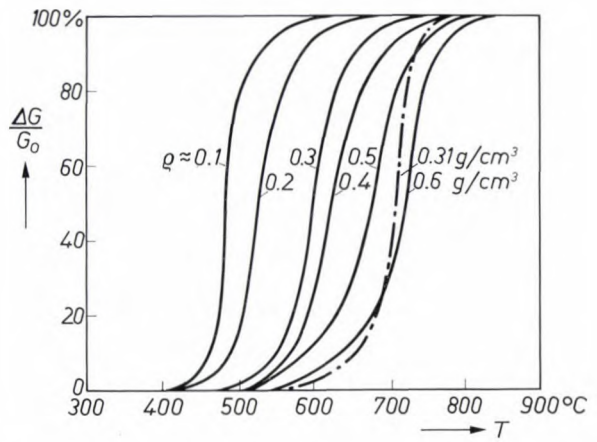
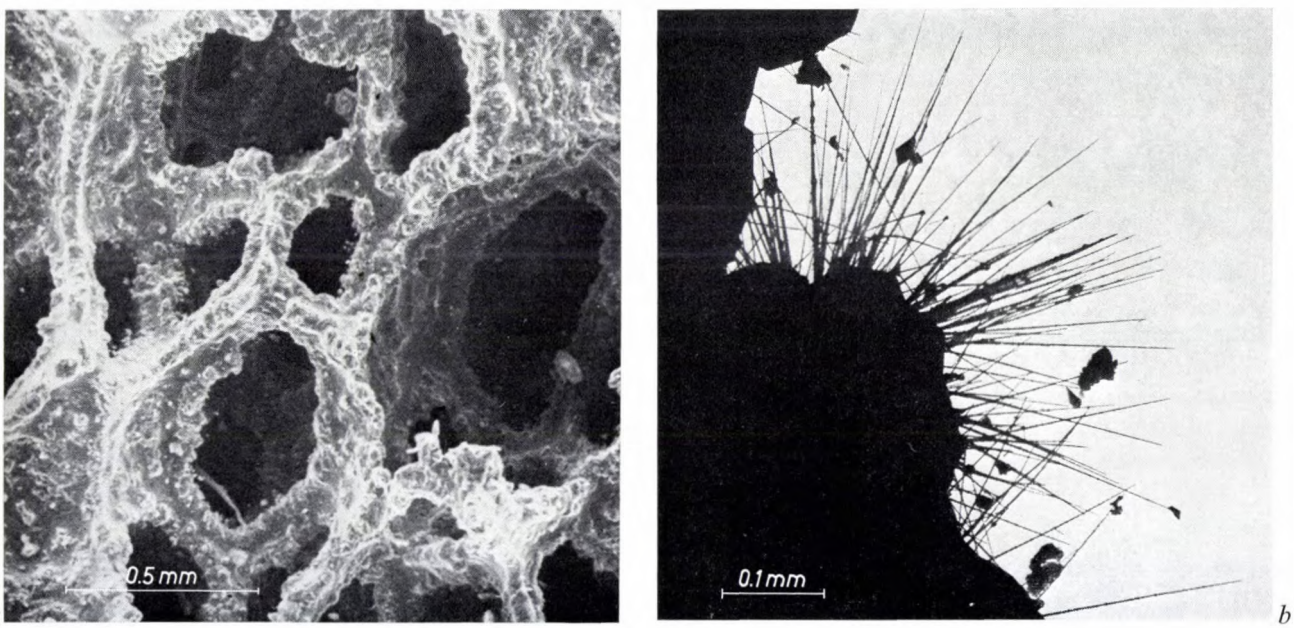


Fig. 14. The mass loss $\Delta G/G_0$ of glassy carbon foam due to oxidation after the material has been left for one hour in a furnace in which the temperature is raised in steps of 50°. The solid curves relate to six types of glassy carbon foam of different density; the chain-dotted lines correspond to a carbon foam partly reinforced with pyrolytic graphite. In this material there is a greater resistance to oxidation, at a relatively low density (0.31 g/cm³), than for glassy carbon foam.

Modifications

Carbon has long been known as a chemically inert substance. If the temperatures are high enough, however, carbon will react with almost any element, whether metal or non-metal. Many carbides have very important technological applications. We are thinking here not so much of the salt-like carbides, such as calcium carbide, but of the metallic carbides, which result from the action of carbon on the elements of the fourth, fifth and sixth subgroups of the periodic table. These refractory carbides, which include the carbides of titanium, zirconium, tantalum, molybdenum and tungsten, represent an important class of materials for machining. Another example is silicon carbide, which is important not only for its hardness and grinding properties but also for its great resistance to oxidation.



a b

Fig. 15. Scanning electron photomicrograph of silicon-carbide foam with a 'smooth' silicon-carbide layer (*a*) and with a silicon-carbide layer covered with whiskers (*b*).

As we have said, glassy carbon foam allows these materials to be manufactured in porous form, with all the associated advantages. In the preparation the appropriate elements for reacting with the carbon can be added to the open pores, usually in the form of gaseous or liquid compounds.

Among the elements that can be combined with glassy carbon foam in this way, tantalum appears to be particularly interesting, since tantalum carbide combines a high melting point (3800 °C) with a low vapour pressure.

In the presence of oxygen, carbon is really no longer a chemically inert substance, even at relatively low temperatures (see fig. 14). An improvement in this respect

is achieved by converting the surface of the glassy carbon foam into silicon carbide by reacting the surface with silicon.

We have tested the resistance to oxidation of this 'silicized' material by exposing it to air for 50 hours at a temperature of 1000 °C. Samples in which the applied silicon had been completely converted to silicon carbide showed no loss in mass under these conditions and no reduction of compressive strength.

This method of 'silicizing' carbon foam can also be carried out in such a way that instead of a 'smooth' silicon-carbide layer (*fig. 15a*) a covering layer is formed which has SiC whiskers that penetrate into the pores (fig. 15*b*). A whiskered skeleton of this type seems to lend itself excellently as a means of reinforcing metals to provide great stiffness.

Some actual and potential applications

The lightness, strength and chemical resistance of the glassy carbon foam we have prepared at once suggest the possibility of using it as a structural material. When we consider its low thermal conductivity, it becomes clear why the material is able to meet the space-technology specifications for self-supporting heat shields (e.g. in the 'Space Shuttle' project). The material is of course generally attractive for heat insulation in the more earth-bound applications as well. The nature of its porosity (which makes the internal surface readily accessible) means that the material could also find useful application as catalyst substrates, corrosion-resistant filters, heat exchangers, etc. Its electrical conductivity allows it to be used in all kinds of electrodes, in combination with the other applications. If its electrical conductivity is suitably adjusted, the material can also be made suitable for the manufacture of heating elements. The possibilities we have described for varying and modifying this material greatly increase its potential applications in industry under a wide variety of conditions.

The new material is also of interest in medical science. Since it is immunobiologically neutral and also, because of its pore structure, provides a good basis on which new tissue can grow, it would seem to be very suitable for internal application in prostheses. It can be given a close functional resemblance to real bones by filling moulds of solid glassy carbon with glassy carbon foam.

Our main application for the material so far is in glass technology. Glassy carbon is already widely used here since it is not 'wetted' by molten glass and it stands up well to the thermal shocks often encountered in glass treatment. The extension of the development to give glassy carbon foam has very greatly widened the useful scope of the material in this field.

Summary. A light and porous type of carbon is described, called glassy carbon foam, which has relatively low thermal and electrical conductivities but is very strong. It is prepared from foamed polyurethane with a reticular structure; the foamed material is successively stabilized, impregnated, e.g. with a phenol-formaldehyde-resin solution, and subjected to a mild pyrolytic treatment. The material owes its strength to the formation of glassy carbon from the resin phase (this is why the pyrolysis is subject to certain restrictions). The density of the material, and hence its compressive and tensile strengths and its thermal and electrical conductivities are controlled by varying the porosity of the original foamed material, the degree of impregnation and the method of pyrolysis. The magnitude and accessibility of the internal surface also allow the material to be converted into a variety of porous metal-carbon composites and carbides. The product obtained when the material is reacted with silicon is highly resistant to oxidation. Glassy carbon foam — and its various modifications and combinations — can be used as a structural material, for heat shields, for corrosion-resistant filters, heat exchangers, electrodes, heating elements, internal prostheses, etc. and in glass technology.

The subject is introduced with a general discussion of the structure and growth of carbon.

Continuous cooling in the millikelvin range

F. A. Staas

The cooling of a system to very low temperatures is a means of investigating weak interactions that would be completely swamped by other effects at higher temperatures. In the temperature range of about 1 to 10 K, superconductivity and superfluidity are among the striking effects that arise because of the dominance of certain interactions. There are quite a number of interesting interactions, however, such as magnetic interactions between nuclear spins, that only become dominant in the millikelvin range. As long ago as the thirties the millikelvin range was reached by means of adiabatic demagnetization, but this method has the disadvantage that the cooling process is a 'single-shot' method. Continuous cooling in this range first became feasible with the advent of the 'dilution refrigerator' in the sixties. The dilution refrigerator described in this article is capable of extracting about ten times more heat per second from the environment than earlier types.

Low-temperature research would nowadays scarcely be conceivable without liquid helium as a 'cold bath'. Under normal pressure helium boils at 4.2 K. In 1908, on the very same day that he became the first to liquefy helium, Kamerlingh Onnes obtained a temperature of 1 K by vigorously pumping away the helium vapour above the liquid (the experiment was in fact part of an abortive attempt to solidify the helium).

Temperatures much lower than 1 K cannot be reached in this way, because the *cooling capacity* then approaches zero. This quantity is determined by the amount of helium vapour that can be removed per second, in other words it depends on the pumping capacity and the vapour pressure. At 1 K the vapour pressure of helium is still 16 Pa (0.12 mm Hg), but at 0.5 K it is no more than 2.1×10^{-3} Pa (1.6×10^{-5} mm Hg). In 1932 W. H. Keesom obtained a temperature of 0.71 K (vapour pressure 0.5 Pa) with a pumping capacity of 675 l/s. These figures relate to naturally occurring helium, which consists almost entirely of ^{4}He. It is possible with the (very expensive) isotope ^{3}He, which has a higher vapour pressure at the same temperature (boiling point 3.2 K), to reach a temperature of about 0.3 K by pumping.

The millikelvin range was opened up in the thirties by means of 'adiabatic demagnetization'. This method, which will not be dealt with here, has the disadvantage for the experimenter that the cooling process occurs once only (single-shot experiment). Continuous cooling in the millikelvin range did not in fact become possible until the sixties, with the advent of the dilution refrigerator discussed in this article.

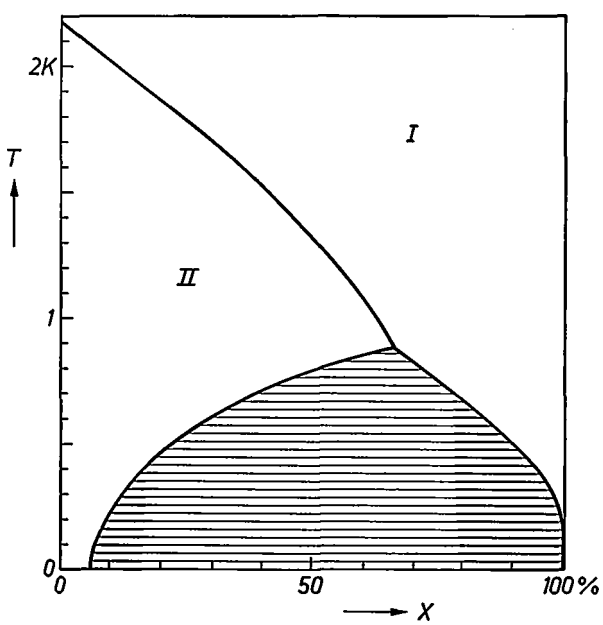


Fig. 1. Phase diagram of liquid mixtures of ^{3}He and ^{4}He. T is the temperature, X the molar concentration of ^{3}He in the mixture; $X = n_3/(n_3 + n_4)$, where n_3 and n_4 are the number of moles of ^{3}He and ^{4}He respectively. In region *I* (He I) the liquid is normal, in region *II* (He II) the ^{4}He component shows superfluid properties. In the hatched region phase separation takes place.

The operation of the dilution refrigerator is based on some special properties of liquid mixtures of ^{4}He and ^{3}He. These are summarized in the phase diagram given in *fig. 1*, which shows the temperature T plotted against the molar concentration X of the ^{3}He in the mix-

Dr F. A. Staas is with Philips Research Laboratories, Eindhoven.

ture. On cooling, pure ^{4}He ($X = 0$) changes at the 'λ point' (2.17 K) from a normal liquid ('He I') into a liquid with superfluid properties ('He II'). Above 0.88 K ^{4}He and ^{3}He are miscible in all proportions; the λ-temperature in this region decreases as X increases. Below 0.88 K, however, the liquid separates to form two phases: a 'dilute mixture' (*d*), which has a low content of ^{3}He, and a 'concentrated mixture' (*c*), with a high content of ^{3}He.

The concentrated mixture is the lighter of the two and thus floats on the dilute mixture. The concentrated mixture behaves as a normal liquid. In the dilute solution, however, the ^{4}He is superfluid, and the ^{3}He atoms contained in it are not impeded in their movement; it is as if for these atoms the ^{4}He does not exist. The ^{4}He thus behaves like a kind of vacuum, and the ^{3}He dissolved in it may be regarded as a gas.

This 'gas' has an exceptionally high density compared with that of a normal gas at the same temperature. In about 1960 the location of the part of the phase-separation line between 0.88 K and 0.4 K was known experimentally. It showed that at 0.5 K, for example, phase *c* is in equilibrium with a phase *d* that contains approximately 20% of ^{3}He, i.e. 7×10^{-3} mol/cm^3. This is more than a thousand times the density of normal ^{3}He gas that is in equilibrium with liquid ^{3}He at 0.5 K. Since these large quantities can indeed be removed selectively from the dilute solution, very effective cooling should be possible. This is the basic principle of the dilution refrigerator [1]. The heart of the refrigerator is the mixing chamber in *fig. 2*. When the ^{3}He is removed from the dilute solution *d*, the ^{3}He in *c* goes over to *d*. Like evaporation, this mixing process takes up heat, since it is the fastest atoms from *c* that have gone over to *d*. This heat is extracted from the chamber, which is therefore cooled.

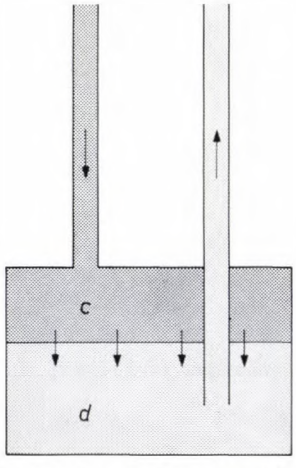


Fig. 2. The mixing chamber, the heart of a dilution refrigerator. When the ^{3}He is extracted from the dilute solution (*d*), ^{3}He goes from the concentrated mixture (*c*) to *d*. This mixing process closely resembles evaporation and, like evaporation, it takes up heat.

The first dilution refrigerators [2] worked *better* than expected. This was because the solubility of ^{3}He in ^{4}He, extrapolated to $T = 0$, does not occur at $X = 0$, as had originally been thought, but at $X = 6.4\%$ [3] (see fig. 1). Unlike the situation with normal vapour, there remains in the dilute solution, even at the lowest temperatures, a considerable quantity of ^{3}He which can be removed.

Dilution refrigerators are now widely used in laboratories all over the world. In most of them, ^{3}He is circulated by means of a pump. The most important of the quantities that determine the cooling capacity is the circulation rate $\dot{n}_3$, i.e. the number of moles of ^{3}He passing through a given cross-section per second. There is also a version in which ^{4}He is circulated instead of ^{3}He [4]; in this type the ^{4}He is conducted towards the phase-separation plane instead of ^{3}He.

In this article we shall discuss a new type of dilution refrigerator, in which ^{3}He and ^{4}He *both* circulate. In this refrigerator we have been able to circumvent some of the typical limitations of the 'conventional' type employing ^{3}He circulation. The result is a cooling capacity one or two orders of magnitude higher than in the conventional refrigerator. By way of introduction we shall first discuss the conventional refrigerator and its limitations, and then very briefly describe the refrigerator using only ^{4}He circulation.

The conventional dilution refrigerator

The operation of the conventional dilution refrigerator is illustrated by the diagram in *fig. 3*. A ^{3}He-^{4}He mixture is cooled to about 1.5 K and then cooled further by pumping. The composition is chosen in such a way that phase separation finally takes place in the mixing chamber *M*. A heater wire *H* may be used to ensure that the temperature of the vaporization chamber (or 'still') *V*, which is connected to *M* via the pipe P_2, does not fall below 0.7 K. At this temperature the partial vapour pressure of ^{3}He above the mixture is still high enough (9 Pa $= 7 \times 10^{-2}$ mm Hg) for the ^{3}He to be effectively pumped away. The ^{4}He evaporates to a much lower extent; the gas contains 98% of ^{3}He. The ^{3}He is thus selectively removed from the dilute solution (*d*), so that heat is extracted from the system in *M* as described in the caption of fig. 2. Also because the concentrated

[1] H. London, G. R. Clarke and E. Mendoza, Phys. Rev. **128**, 1992, 1962.
[2] H. E. Hall, in: J. F. Allen (ed.), Superfluid helium, Academic Press, London 1966, p. 7.
[3] D. O. Edwards, D. F. Brewer, P. Seligman, M. Skertic and M. Yaqub, Phys. Rev. Letters **15**, 773, 1965 (predicts $X = 5\frac{1}{2}\%$).
[4] K. W. Taconis, N. H. Pennings, P. Das and R. de Bruyn Ouboter, Physica **56**, 168, 1971.
N. H. Pennings, K. W. Taconis and R. de Bruyn Ouboter, Cryogenics **14**, 53, 1974.

^{3}He (*c*) entering through the pipe P_1 via the heat exchanger W is precooled by the liquid in the pipe P_2, the temperature in M falls to 0.1 K or lower. The liquid in P_2 is preheated in W. The ^{3}He evaporating in V is returned at a pressure of 13×10^3 Pa (10 cm Hg) by an external pump to the chamber C, which is kept at 1.5 K by thermal contact with an evacuated helium bath, where the gas condenses. The concentrated ^{3}He returns to P_1 through the flow resistance W', which takes up the pressure difference of 13×10^3 Pa between C and P_1 and is in thermal contact with V. The diagram in *fig. 4* indicates the phases through which the liquid passes in going from C to V.

The part of the system drawn in fig. 3 is contained in a vacuum. The only connections with the outside world are the pump leads, the thermal contact of C with the helium bath, the heater H and other measurement and control leads.

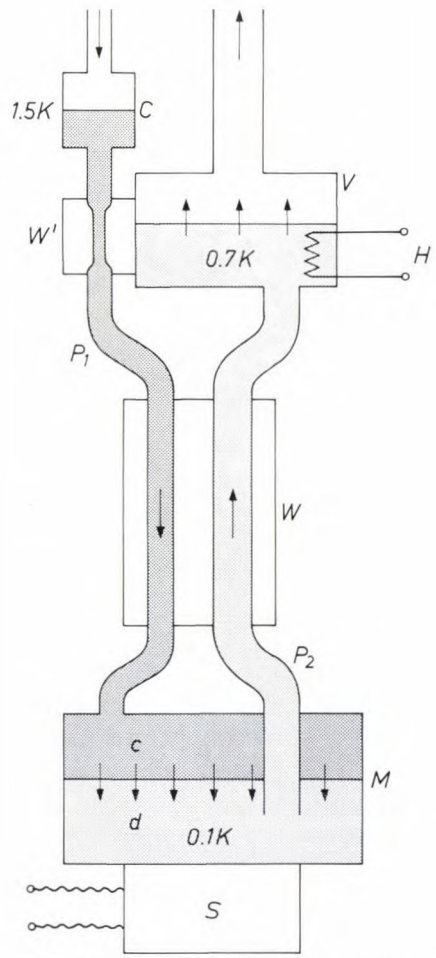


Fig. 3. A conventional dilution refrigerator. Cooling takes place in the mixing chamber M, where ^{3}He of the concentrated mixture (*c*) transfers to the dilute solution (*d*). The ^{3}He passes through the pipe P_2 and enters the vaporization chamber (or still) V, from which it is pumped away; the vapour consists almost entirely of ^{3}He owing to the great difference in the vapour pressures of ^{3}He and ^{4}He. The ^{3}He is compressed to about 10 cm Hg, condenses in the vessel C (1.5 K) and returns to the mixing chamber via the heat exchanger and flow resistance W' and the pipe P_1. W heat exchanger. S sample. H heater.

Fig. 5 shows a conventional dilution refrigerator developed at Philips Research Laboratories. Six models of this type have been operated in various laboratories by way of trial [5]. Particular attention has been paid to the heat exchangers in these very compact machines: these heat exchangers present a problem which we shall examine more closely later. Another difference compared with many other dilution refrigerators is that the flow resistance W' is variable and pressure-dependent; it opens up as the pressure increases. This makes it easier to work with the refrigerator; starting in particular is faster. In the new refrigerator, which will be discussed later, the rate of ^{3}He circulation is much larger, so that a pressure control of this type is a practical necessity.

Just as when pumping out an ordinary gas, the pipe P_2 should not form a high 'flow resistance'. If it did, the concentration of ^{3}He contained in the liquid in the vaporization chamber would be too low, and so too therefore would the vapour pressure above the liquid. To prevent this, P_2 must be made sufficiently wide. In this problem of ^{3}He transport through P_2 the concept of the 'pressure' of the ^{3}He in the dilute solution plays an important part. It must not, for example, have any significant gradient. We shall now take a closer look at this concept.

The dilute solution

The pressure to be assigned to the ^{3}He 'gas' in the dilute solution is in fact the *osmotic pressure* of ^{3}He in superfluid ^{4}He. The reality of this can be demonstrated in an arrangement like that in *fig. 6*. Two vessels are

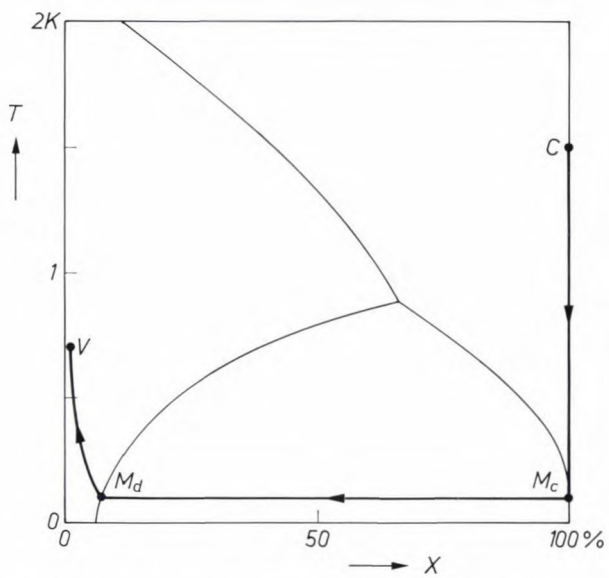


Fig 4. Phase diagram showing the phases which the liquid passes through between C and V in fig. 3. M_c is the concentrated phase in M, M_d the dilute phase.

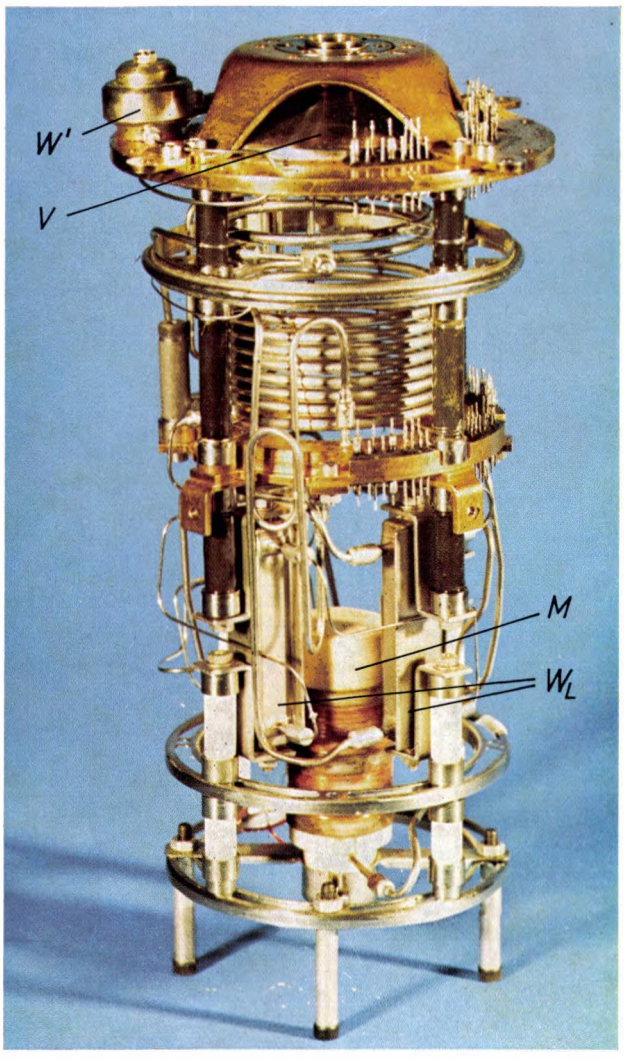


Fig. 5. A conventional dilution refrigerator developed at Philips Research Laboratories. *M* mixing chamber. *V* vaporization chamber (or still). W_L the last heat-exchanger sections before the mixing chamber, each with a foil partition of 350 cm². *W'* heat exchanger and pressure-dependent flow resistance.

mixing chamber decreased. In fact, however, the situation is much more favourable: even at $T = 0$ the osmotic pressure of the equilibrium solution of 6.4% ^{3}He in ^{4}He is still 2300 Pa (17 mm Hg). The pressure has this value because the ^{3}He behaves like a Fermi-Dirac gas, as we shall now briefly explain.

In quantum-statistical terms, there are two classes of particles: bosons and fermions. ^{4}He belongs to the first, ^{3}He to the second. If Bose-Einstein statistics are applicable, each quantum state can be occupied by an arbitrary number of particles, whereas in Fermi-Dirac

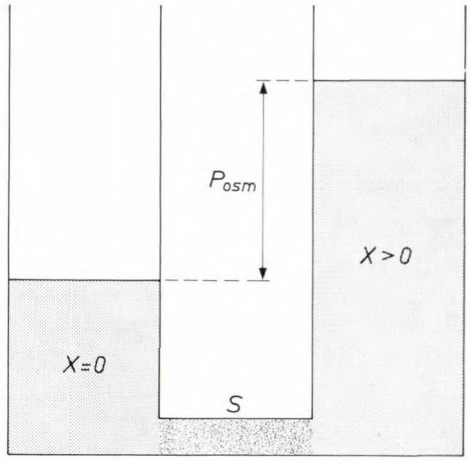


Fig. 6. Osmosis of ^{3}He in ^{4}He. The left-hand vessel contains pure ^{4}He, the right-hand vessel a mixture of ^{3}He and ^{4}He. The superleak *S* passes the superfluid ^{4}He but not the ^{3}He. The difference in level in millimetres, P_{osm}, is the osmotic pressure measured in millimetres of helium. The ^{3}He may be thought of as a gas contained within the volume of the ^{4}He in the right-hand vessel. The surface of the liquid constitutes an impermeable membrane for this 'gas', owing to the strength of the Van der Waals forces that prevent it from passing through the surface. The gas tries to expand, and such expansion can take place by displacement of the liquid through the superleak. This displacement continues until the pressure of the gas (the osmotic pressure) is in equilibrium with the hydrostatic pressure. A similar explanation would apply to the osmosis of sugar in water.

connected by means of a 'superleak'. This is a tube filled with a powder so fine that only superfluid ^{4}He will pass through it. The left-hand vessel contains superfluid ^{4}He, the right-hand one a mixture of superfluid ^{4}He with ^{3}He. As in the case of water and a sugar solution separated by a semipermeable membrane, the osmotic pressure in the arrangement shown in fig. 6 gives rise to a difference of level.

According to Van 't Hoff's equation, sugar in water behaves as an ideal gas: the osmotic pressure is

$$P = \varrho RT, \qquad (1)$$

where ϱ is the molar density of the sugar. If this relation were also to apply to ^{3}He in ^{4}He, circulation would eventually become impossible because the pressure of the ^{3}He would fall to zero as the temperature in the statistics a quantum state can be occupied by no more than one particle. As an illustration, *fig. 7* shows the distribution at $T = 0$ for a number of particles over the quantum states in the momentum space of a free particle in two dimensions. The bosons (fig. 7*a*) are all in the lowest state (Bose-Einstein condensation), whereas the fermions (fig. 7*b*) fill a 'sphere' in momentum space. The bosons (^{4}He) need not be considered here. We merely note that superfluidity is usually attributed to Bose-Einstein condensation.

Even at $T = 0$, fermions therefore possess momentum and kinetic energy, and hence also exert a pressure at $T = 0$. This pressure is generally given by:

$$P = \tfrac{2}{3} U, \qquad (2)$$

[5] These were built by A. P. Severijns of these laboratories.

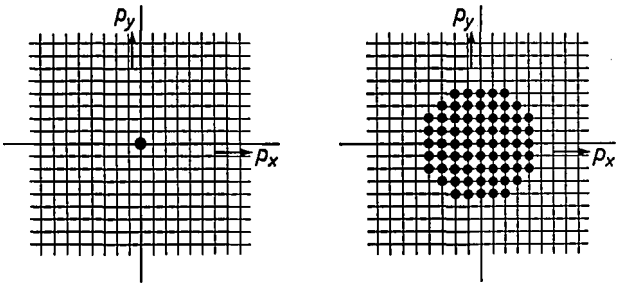


Fig. 7. Momentum space for a free particle in two dimensions. The lattice points are the quantum states. At $T = 0$ all particles condense in the lowest state in the case of Bose-Einstein statistics (*a*). In the case of Fermi-Dirac statistics the particles at $T = 0$ occupy a 'sphere' of lowest states (*b*). Thus, at $T = 0$, Fermi particles (fermions) still possess an average momentum. This accounts for the osmotic pressure which ^{3}He in ^{4}He still possesses at $T = 0$.

where U is the kinetic energy per cm^3. (The ideal gas law (1) follows from (2) with $U = \frac{3}{2} \varrho RT$ for a Boltzmann gas.) We consider a volume of 1 cm^3 and fill it at $T = 0$ with one particle after another. The radius p of the Fermi sphere — the sphere of occupied states in momentum space — then steadily increases. The number of particles present at a given instant in the volume of 1 cm^3, i.e. the density n, is proportional to the volume of the Fermi sphere, i.e. to p^3, if n is large. On the other hand, the energy ε of the last particle added is equal to $p^2/2m$, where m is the mass per particle. Consequently ε is proportional to $n^{2/3}$. *Fig. 8* gives a plot of ε as a function of n. The total energy U is the sum of all the particle energies. For large n the quantity U can be expressed as an integral equal for $n = n_0$ to the cross-hatched area in fig. 8. The effective mass of ^{3}He atoms in superfluid ^{4}He is 2.34 times the true mass. Using this to calculate the integral to a limit n_0 corresponding to $X = 6.4\%$, we then find from (2) a pressure of 2900 Pa (22 mm Hg) for the Fermi-Dirac gas. The discrepancy from the experimental value (2300 Pa) must be attributed to a weak interaction between the ^{3}He atoms.

Fig. 9 demonstrates once again the practical importance for the dilution refrigerator of the fact that the ^{3}He behaves like a Fermi-Dirac gas.

Finally, we shall briefly consider the solubility of ^{3}He in ^{4}He. The volume of liquid ^{3}He is 37 cm^3 per mole, that of ^{4}He is 28 cm^3 per mole. Thus, ^{3}He atoms occupy a greater volume in the liquid than ^{4}He atoms. Theoretically this would be expected: the amplitude of the zero-point vibration, which each enclosed particle still performs at $T = 0$, is greater for the lighter ^{3}He atoms. *Fig. 10* pictures the situation for liquid ^{3}He floating on liquid ^{4}He; the difference in atomic volumes is greatly exaggerated. Owing to the attractive forces (Van der Waals forces) the particles tend to bunch as closely as possible together. A ^{4}He atom will not tend to move towards the ^{3}He, where it would have fewer neighbours at a greater distance (measured from centre to centre). Conversely, a ^{3}He atom will prefer to move towards the ^{4}He, since it will then have more neighbours at a shorter distance. On the other hand, the ^{4}He atoms will have to move away from each other to admit a ^{3}He atom. As confirmed by theoretical calculations [6] it turns out that energy is gained upon such a transition, i.e. the *thermodynamic potential* μ_c of ^{3}He atoms in pure liquid ^{3}He in relation to that in pure liquid ^{4}He is positive. As the Fermi sphere fills up, however, the energy gain decreases with the increase in the number of ^{3}He atoms in the ^{4}He, and equilibrium is reached at $X = 6.4\%$. In fig. 8, ε may be regarded as the thermodynamic potential μ_d of ^{3}He atoms in a dilute solution

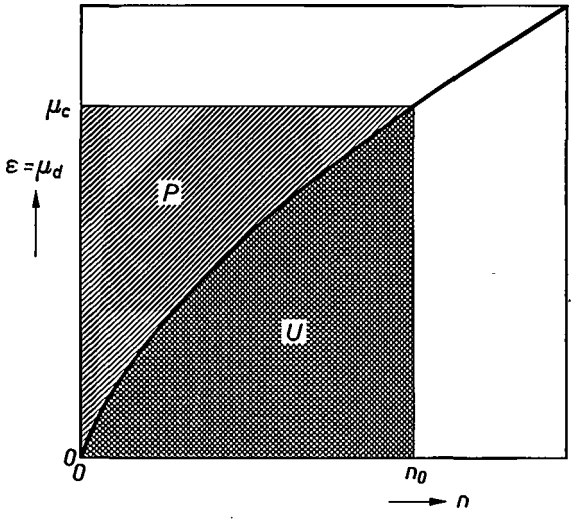


Fig. 8. The energy ε of a particle as a function of the particle number n in the case of Fermi-Dirac statistics when a volume of 1 cm^3 is filled by particle after particle at $T = 0$. The energy per cm^3 (U) of a Fermi-Dirac gas with n_0 particles is equal to the cross-hatched area. From the relation $\varepsilon \propto n^{2/3}$ it follows that the hatched area is equal to 2/3 of the cross-hatched area. This is therefore equal to the pressure P. Applying this approach now to helium, ε is the thermodynamic potential μ_d of ^{3}He atoms in the dilute solution (compared with that in pure ^{4}He). A dilute solution is in equilibrium with concentrated ^{3}He when μ_d is equal to the thermodynamic potential μ_c of ^{3}He atoms in the concentrated ^{3}He.

with respect to that in pure ^{4}He. At equilibrium, μ_d is equal to μ_c.

The energy gain on the transition of a ^{3}He atom from pure ^{3}He to pure ^{4}He is in sharp contrast with the considerable energy loss, due to the Van der Waals forces in the liquid, that occurs on the transition of an atom from the liquid to true vacuum. It is of course the magnitude of this energy loss that accounts for the vapour pressure being so small at low temperature. In ^{4}He the Van der Waals forces are stronger than in ^{3}He and therefore the vapour pressure of ^{4}He is lower at the same temperature, as we have already seen.

Cooling capacity and lowest attainable temperature of dilution refrigerators

Heat exchangers

One of the factors that limits the cooling capacity of the conventional dilution refrigerator is the heat exchanger W in fig. 3. Why should this be?

The cooling takes place because the ^{3}He atoms in the dilute solution possess more energy than those in the concentrated mixture. Furthermore, the ^{3}He in the dilute solution in the pipe P_2 takes up additional heat from the concentrated ^{3}He in P_1 because it expands (P_{osm} is constant, T increases and X decreases; see fig. 9). The internal energies in the two phases are known from measurements of specific heats. Below 0.7 K the energy transported downwards through P_1 by the concentrated solution per mole of ^{3}He is αT^2, and the energy transported upwards by the diluted solution in P_2 is given by βT^2 per mole of ^{3}He, where α is equal to 12 J/mol K^2 and β to 94 J/mol K^2.

The energy-flow balance over one of the closed surfaces a, b or c in *fig. 11* is therefore given by:

$$Q + \dot{n}_3 \alpha T_1^2 = \dot{n}_3 \beta T_2^2. \qquad (3)$$

Q is the heat absorbed per second from the environment by the mixing chamber, i.e. the cooling capacity. T_1 and T_2 are the temperatures of the liquids in P_1 and P_2 at the level of the chosen cross-section.

A complete analysis leads to rather complicated calculations [7], but the following simple considerations give a qualitative picture of the situation. The temperature in the lower part of P_2 is equal to the temperature T_M of the mixing chamber, $T_2(a) = T_M$. Let us now first assume that the heat exchanger is so good that the temperatures at the lower end are identical: $T_1(a) = T_2(a) = T_M$. From the balance over the area (a) it then follows that:

$$Q = \dot{n}_3(\beta - \alpha)T_M^2 = 82\,\dot{n}_3 T_M^2\,. \qquad (4)$$

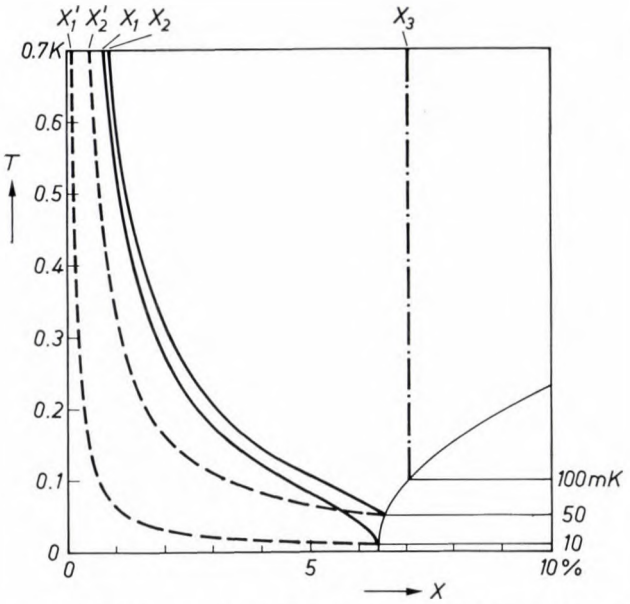


Fig. 9. Left-hand part of the phase diagram, with lines of constant osmotic pressure (solid lines), for a dilute solution which is in equilibrium with concentrated ^{3}He at 50 mK and 10 mK. These lines are followed in the pipe P_2 of fig. 3 if the pressure drop in the pipe is negligible. If the ^{3}He behaved as an ideal gas and not as a Fermi-Dirac gas, the dashed lines would be followed (XT constant, see eq. 1), and the concentration in the still would have been much smaller, i.e. X_1' and X_2' instead of X_1 and X_2. The vertical line (X constant) from a concentrated mixture of 0.1 K relates to the new refrigerator (p. 111); this gives a much greater concentration (X_3) in the vaporization chamber.

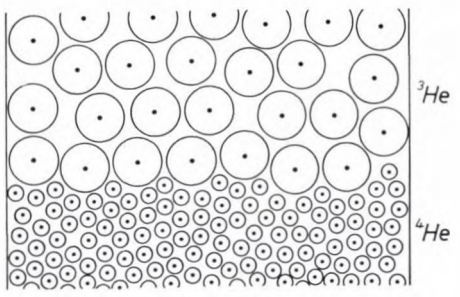


Fig. 10. Simple representation of liquid ^{3}He floating on liquid ^{4}He. The difference in atomic volumes is greatly exaggerated.

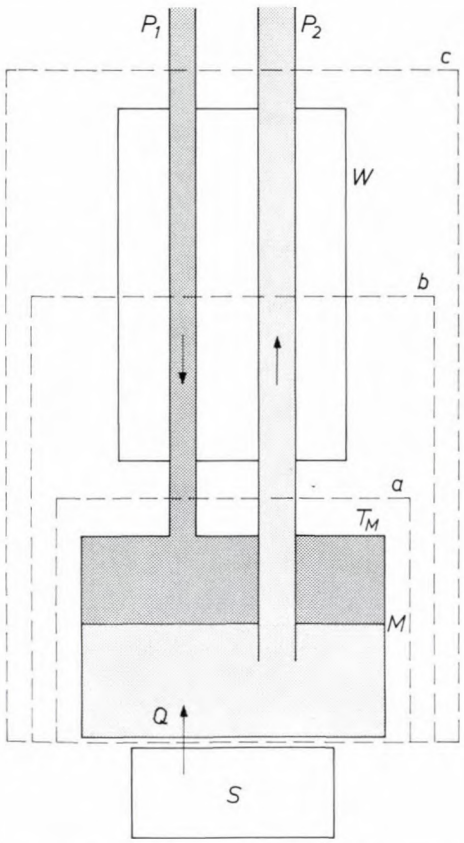


Fig. 11. Lower part of the conventional dilution refrigerator, with closed surfaces a, b and c over which the energy balance is taken. M, W, S, P_1, P_2 are as in fig. 3.

[6] L. J. Campbell, Phys. Rev. Letters **19**, 156, 1967.
[7] F. A. Staas, K. Weiss and A. P. Severijns, Cryogenics **14**, 253, 1974.

Eq. (4) gives the cooling capacity Q at a given T_M, or the lowest attainable temperature T_M for a given heat load Q. The equation shows that in the first case the cooling capacity increases, and in the second case the attainable temperature decreases for increasing values of $\dot{n}_3$.

Raising $\dot{n}_3$ indefinitely, even if it could be done, is however no help if the heat exchanger is not ideal. This is because only a limited quantity of heat can be transported per unit time from P_1 to P_2. With increasing $\dot{n}_3$ a value is eventually reached at which it is no longer possible to cool the liquid in P_1 from 0.7 K, the starting temperature, to the temperature of the mixing chamber. It then follows at once from eq. (3) that the cooling capacity at a given $\dot{n}_3$ is smaller than it was before, since T_1 is now greater than T_2 at the lower end of W. Let us now suppose, however, as a second extreme case, that the heat leakage Q is zero. Then from eq. (3), for each cross-section between a and c:

$$T_2 = \sqrt{\alpha/\beta}\, T_1 = 0.35\, T_1. \qquad (5)$$

The lowest attainable temperature is now determined by the lowest value of $T_1(a)$ that can be reached with the heat exchanger:

$$T_M = T_2(a) = 0.35\, T_1(a). \qquad (6)$$

Taken together this yields the following picture. If we let the circulation rate $\dot{n}_3$ increase from zero, the heat exchanger is initially adequate, and eq. (4) is approximately valid. This implies that at a given Q the temperature T_M decreases. At a later stage, however, the heat exchanger is no longer able to cope. $T_1(a)$ then begins to rise. Since at the same time Q has then become relatively small, eq. (6) is approximately valid, so that T_M also starts to rise. There is therefore an optimum value for $\dot{n}_3$, which is higher the better the heat exchanger.

In a good heat exchanger the two liquid flows are separated by only a thin foil with the largest possible surface area. Heat exchangers of this type serve their purpose well down to about 100 mK. As the temperature decreases further, however, they rapidly deteriorate because of the marked increase in the Kapitza resistance. This is the resistance to the heat flow that arises at the interface between two dissimilar media, and is a consequence of the difference in the velocity of sound in the two media (due to differences in density and stiffness). Owing to this acoustic mismatch the majority of the phonons — i.e. the waves that are responsible for the heat transport — are reflected from the surface. In a heat exchanger for liquid helium the Kapitza resistance is about 10^4 times as large as the thermal resistance of the metal of the wall itself. The steep rise with falling temperature is attributable to a decrease in the phonon density. The heat flow Q_K through the interface can be represented by

$$Q_K \propto T_1^4 - T_2^4. \qquad (7)$$

If there is any significant difference between T_1 and T_2, as in the case where the heat leakage is negligible (see eq. 5), then Q_K is almost completely determined by T_1. Equation (7) shows that Q_K is strongly dependent on T_1.

The increase in the Kapitza resistance is a serious obstacle to the attainment of temperatures below 10 mK. An added difficulty is that the viscosity of liquid ^{3}He behaves very differently from that of the superfluid: it *increases* with falling temperature. Because of the large surface area necessary in the heat exchanger, at low T and high $\dot{n}_3$ this can easily lead to considerable viscous heat generation in P_1, which counteracts the cooling.

Largely because of the improvements in heat exchangers there are now conventional dilution refrigerators available that can reach a $T_1(a)$ of about 30 mK, and therefore a T_M of about 10 mK. In the refrigerator shown in fig. 5 the part of the heat exchanger at the lowest temperature is divided into four sections W_L; the partition in each of these, separating the liquids, is a folded sheet of foil with a surface area of 350 cm^2.

The optimum value of $\dot{n}_3$ determined by the heat exchanger amounts to about 10 μmol/s in small refrigerators, and to about 100 μmol/s in large ones. This corresponds to cooling capacities at 100 mK of the order of 10 μW and 100 μW respectively, as may be deduced from eq. (4). In the new type of dilution refrigerator to be discussed shortly, the heat exchanger problem has been completely bypassed. As a result it is advantageous to use higher rates of circulation. With an $\dot{n}_3$ of about 1 mmol/s we have achieved a cooling capacity at 100 mK of the order of 1 mW.

Other factors

In the still of the conventional dilution refrigerator the concentration of the ^{3}He in the liquid, and consequently the vapour pressure of the ^{3}He above the liquid, is fairly low (see fig. 9). This implies that to reach a large $\dot{n}_3$ a high pumping capacity is required. The very high circulation rate of 1 mmol/s used in the new refrigerator would necessitate a very expensive pump installation in a conventional machine. As we shall show, this is not the case in the new refrigerator because we have succeeded in applying a special method of forced circulation of the ^{3}He in the dilute solution, increasing the vapour pressure in the vaporization chamber by about 30 times.

Finally, we should mention the following difficulty in the conventional refrigerator. At the top of the pipe P_2

the temperature is higher and X therefore lower (see fig. 9), which means that the liquid at the top is heavier than at the bottom. Convection can thus easily be caused by gravity, resulting in a higher ^{3}He concentration near the top. This means that the ^{3}He 'gas' is not able to expand as much, and thus takes up less heat from the inflowing concentrated ^{3}He. In the conventional cooler shown in fig. 5, convection is avoided by arranging the four sections W_L of the heat exchanger in such a way that in these sections the ^{3}He in the dilute phase flows downwards. As will be seen from the following, the convection problem is not encountered in the new dilution refrigerator.

Cooling with ^{4}He circulation

The principle of cooling by ^{4}He circulation, referred to as 'superfluid injection', is illustrated in *fig. 12*. Two chambers, M_1 and M_2, connected by a relatively wide tube, are filled with liquid helium mixed in such a ratio that when M_2 is cooled a phase separation takes place in it. Owing to the density difference between the two phases, M_1 is then automatically filled with concentrated ^{3}He. Superfluid ^{4}He is now continuously injected into M_1 through the superleak S_1. A new interface is thus produced between the superfluid and the concentrated ^{3}He. Here the mixing takes place that gives the required cooling. The dilute solution continuously produced in this way overflows and 'falls' through the interconnecting pipe into M_2. Whether this does in fact take place in the form of 'droplets' as suggested in fig. 12, we do not know, but it will be convenient here to speak of droplets.

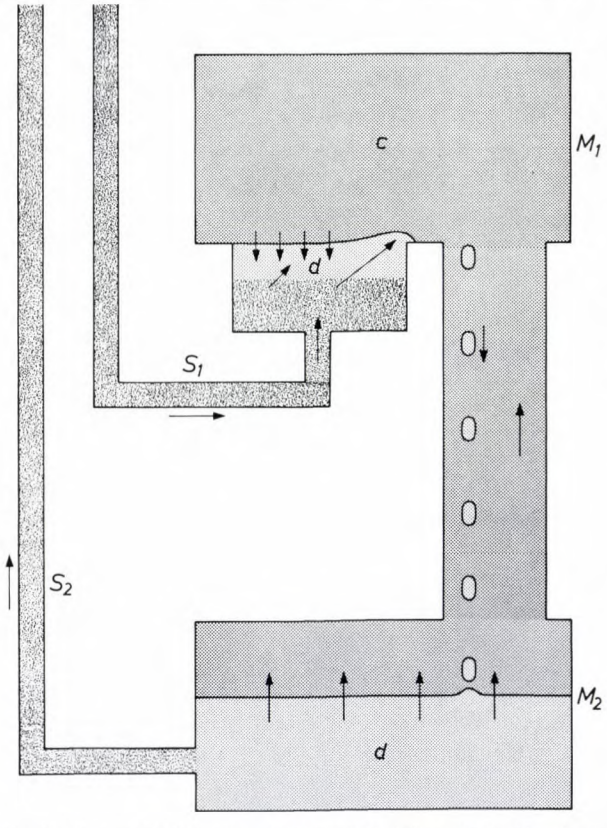


Fig. 12. Superfluid injection. Superfluid is injected into the chamber M_1 through the superleak S_1. The ^{3}He (c) present in this chamber dissolves in it, extracting heat from the environment. The dilute solution produced drops down to the lower chamber M_2. In the refrigerator with ^{4}He circulation [4], superfluid pure ^{4}He is extracted through the superleak S_2; the ^{3}He remaining in M_2 returns to the concentrated mixture. The heat developed in this demixing process is dissipated via a ^{3}He bath under reduced pressure in contact with M_2.

In the Kamerlingh-Onnes laboratory at Leiden a cooling system has been made [4] in which the superfluid ^{4}He is extracted from M_2 through a superleak S_2 and returned to S_1. The ^{3}He supplied by the droplets cannot pass through S_2, and has to return through the phase-separating interface in M_2 to the concentrated mixture c. In this segregation process ('condensation' in the liquid-vapour model) heat is generated which is removed by means of thermal contact between M_2 and an evacuated ^{3}He bath; in this way M_2 is kept at 0.4 K.

The refrigerator thus obtained has two intrinsic advantages: there is no convection problem and it gives almost ideal heat exchange because of the direct thermal contact between the two liquids. A temperature of 10 mK has been reached in M_1, and the circulation rate and cooling capacity are about the same as those of a large conventional refrigerator. Any further increase in the circulation rate, however, is counteracted by Kapitza-resistance problems in the removal of the heat generated in M_2 to the ^{3}He bath.

On the other hand, the liquid in such a refrigerator can be put under pressure, since there is no free liquid surface. At about 33 bars the ^{3}He solidifies, and on a further increase of pressure the solid ^{3}He is cooled further — through an effect that will not be dealt with here — although the cooling is not continuous. In this way a temperature of 3 mK has been reached.

The new refrigerator with ^{3}He and ^{4}He circulation

The dilution refrigerator that we have developed [8] also uses superfluid injection, with the intrinsic advantages mentioned above. Instead of segregation, however, there is mixing in the lower chamber since this itself, acting in turn as a mixing chamber, forms part of a conventional dilution refrigerator. In this way the temperature in M_2 can be kept at 0.1 K, with no Kapitza-resistance problems.

Fig. 13 shows a diagram of the system. M_1 and M_2 are the two chambers of the refrigerator. The conventional arrangement incorporating M_2 differs from that

[8] F. A. Staas, A. P. Severijns and H. C. M. van der Waerden, Physics Letters **53A**, 327, 1975.
See also: G. Frossati, G. Schumacher and D. Thoulouze, Proc. 14th Int. Conf. on Low Temp. Physics, Otaniemi (Finland) 1975, Vol. 4, p. 13.

in fig. 3 in that the superfluid in the pipe P_2 is now not stationary; P_2, together with V and S_2, now forms the return channel for the superfluid that is pumped through S_1 to M_1. This permits an enhanced circulation of ^{3}He, as will presently be shown. First, however, we shall discuss the ^{4}He circulation.

^{4}He circulation

The ^{4}He is circulated by means of a *fountain pump* [9] (F in fig. 13). To explain its operation we shall recapitulate some of the properties of superfluid ^{4}He with reference to the 'two-fluid model'. In this model the liquid is taken as being composed of a *superfluid component* and a *normal component*. The density ϱ_n of the normal component depends on the temperature, as indicated in *fig. 14*. The entropy is carried entirely by the normal component.

The normal component is unable to pass through a superleak. This gives rise to the difference in level in the situation shown in *fig. 15*, which closely resembles that of fig. 6. This effect, known as the fountain effect, can in a sense be regarded as osmosis of the normal into the superfluid component: the superfluid is drawn to the side where ϱ_n is greatest. (Nevertheless the term osmosis is better avoided here, since each component can be converted into the other, which is not usually the case in osmosis.)

A simple fountain pump is illustrated in *fig. 16*. Two chambers, A and B, are interconnected by the capillary C, and connected through the superleaks S_1 and S_2 with the ^{4}He in the vessel V. Chamber B is kept at the

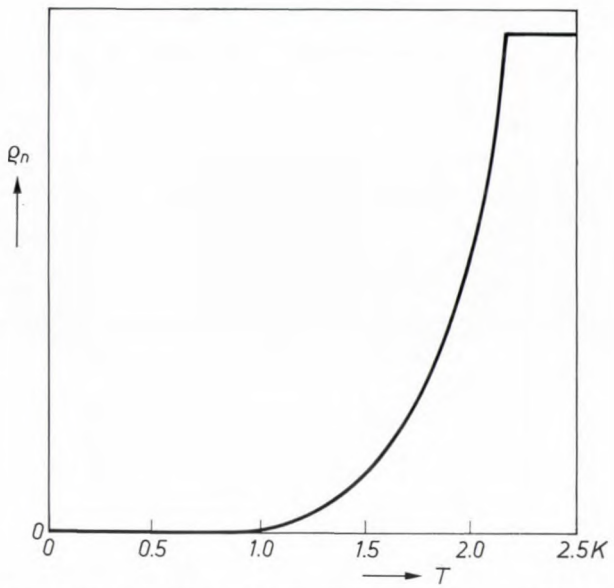


Fig. 14. The density ϱ_n of the normal component in ^{4}He as a function of temperature T.

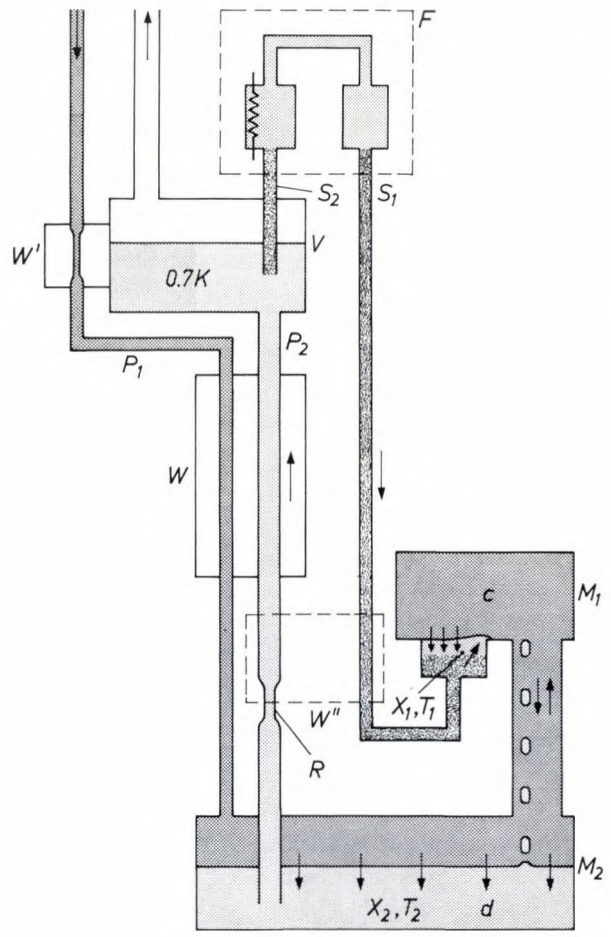


Fig. 13. The new dilution refrigerator with ^{3}He and ^{4}He circulation. In M_1 the cooling takes place as in fig. 12. The lower chamber forms part of a conventional arrangement, as in fig. 3. The pipe P_2, together with the superleak S_2, now forms the return passage for the superfluid injected into M_1, which is circulated by the fountain pump F (see fig. 16). The restrictions (R) in P_2 cause the critical velocity v_{cs} of the superfluid to be greatly exceeded, with the result that the ^{3}He is drawn along, giving a greatly increased ^{3}He circulation.

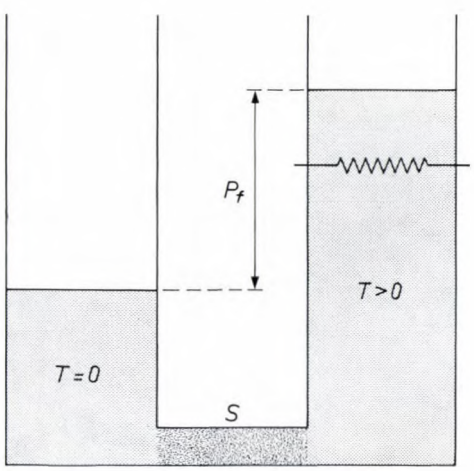


Fig. 15. The fountain effect. A temperature difference between two vessels filled with superfluid ^{4}He and interconnected by a superleak gives rise to a pressure difference. The difference in level is the fountain pressure P_f in the right-hand vessel when the temperature in the left-hand vessel is zero. The effect is attributable to the normal component, and is therefore very small at temperatures below about 1 K (see fig. 14). Apart from its use in the fountain pump, it is therefore not very significant either in the conventional or in the new dilution refrigerator.

temperature T of the bath by heat exchangers; A is kept at a temperature $T + \Delta T$ by means of an additional heat supply.

Since the normal component cannot pass through the superleak, the superleaks S_1 and S_2 transmit no entropy; the heat Q generated in A is therefore completely removed through C and B. This heat transport is equivalent to transport of the normal component, which is formed in A and then converted into a superfluid in B. Because of the fountain effect, superfluid is now drawn from V into A via S_2. The same effect, however, would cause superfluid to flow from B to A via C, if it were not for the fact that the superfluid is dragged along by the normal component that flows in C from A to B. This drag effect occurs when the normal component exceeds a critical speed v_{cn}; the liquid then becomes turbulent. In designing C close tolerances therefore have to be observed. If the capillary is too wide, the heat conduction will be so good that the desired temperature difference ΔT will not be attainable; if it is too narrow, the normal component encounters such resistance that the velocity and hence also $\dot{n}_4$ remain too low.

It is only because of this fountain pump that the refrigerator illustrated in fig. 13 can reach the desired $\dot{n}_4$. Above 1 K the fountain-pressure increases rapidly with temperature; the fountain-pressure difference between 1 K and 1.8 K can be as much as 20 000 Pa (150 mm Hg). A pressure as high as this is necessary to make the system work. This is because the fountain pressure has to oppose the osmosis of the ^{3}He in V, which would try to draw the superfluid through S_2 into V. Another point, as we shall see, is that the pump is 'loaded' by the ^{3}He circulation.

^{3}He circulation

If the ^{4}He liquid flowing through a tube exceeds a critical velocity v_{cs}, the superfluid becomes turbulent. (This is *not* the same effect as the turbulence in tube C in fig. 16; v_{cs} and v_{cn} depend in a different way on the tube diameter.) This property is utilized in the refrigerator to increase the circulation of ^{3}He. This is done by including restrictions in the pipe P_2 in such a way that the liquid flow is increased to a velocity that is a multiple of the critical velocity. As a result, the ^{3}He is drawn along by the ^{4}He, the velocities of the ^{3}He and ^{4}He become identical and $\dot{n}_3$ becomes equal to $X/(1-X)$ times $\dot{n}_4$. The dilute mixture therefore changes very little in composition, as expressed by the vertical line we have drawn in fig. 9. At the temperature of 0.7 K the vapour pressure in V reaches about 270 Pa (2 mm Hg).

Some of the concentrated ^{3}He pumped through P_1 into M_2 (circulation rate $\dot{n}_3$) makes a detour via M_1, mixing there and then entering the dilute solution in M_2 by way of the droplets (internal circulation $\dot{n}_{31}$); the remainder ($\dot{n}_{32}$) mixes directly in M_2. We now have

$$\dot{n}_3/\dot{n}_4 = X_2/(1-X_2),$$

$$\dot{n}_{31}/\dot{n}_4 = X_1/(1-X_1),$$

and therefore

$$\dot{n}_{32}/\dot{n}_4 = X_2/(1-X_2) - X_1/(1-X_1).$$

Here X_1 and X_2 are the concentrations on the phase-

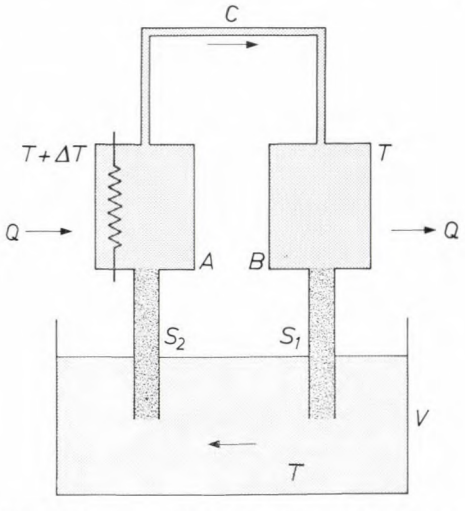


Fig. 16. Fountain pump. When B is kept at the temperature T of the helium in vessel V by thermal contact, and A is raised to a temperature $T + \Delta T$, the ^{4}He is caused to circulate. S_1 and S_2 are superleaks.

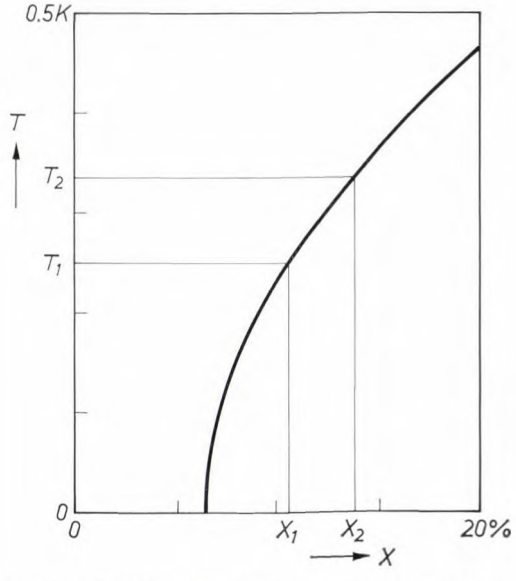


Fig. 17. Detail of the phase diagram, with the temperatures T_1 and T_2 and the concentrations X_1 and X_2 of the dilute solutions in the mixing chambers M_1 and M_2 of fig. 13. When T_1 and T_2 are lower than about 0.1 K, X_1 and X_2 are approximately equal. No further mixing then takes place in M_2, so that this chamber remains at a temperature of about 0.1 K.

[9] F. A. Staas and A. P. Severijns, Cryogenics **9**, 422, 1969.

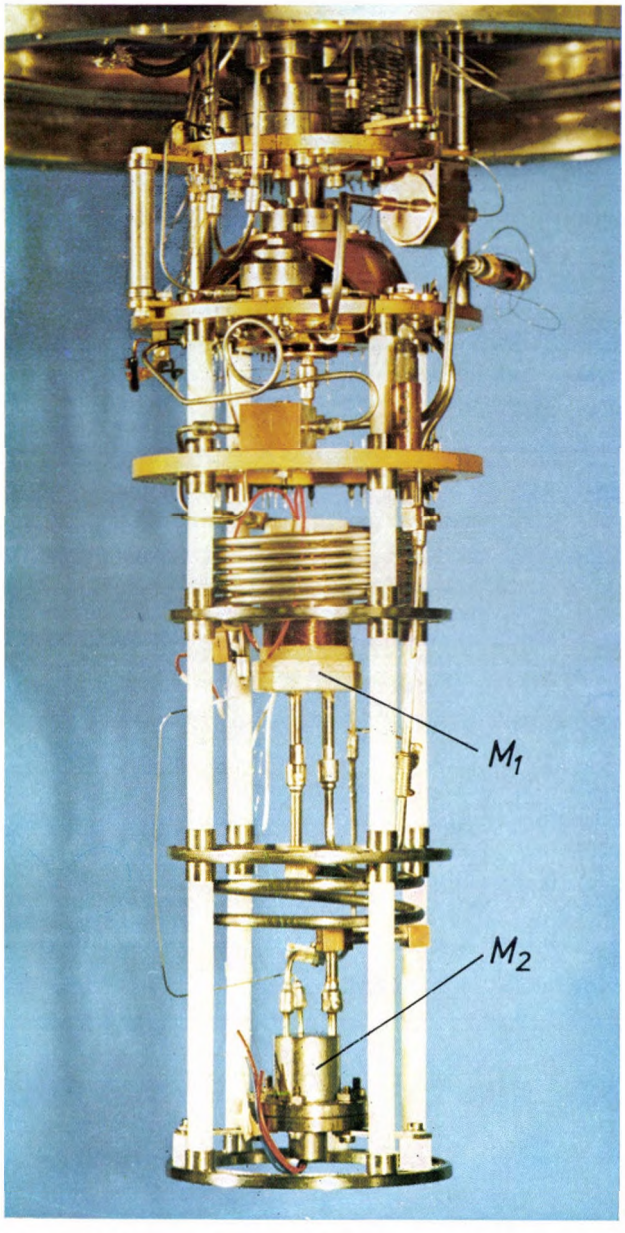


Fig. 18. Experimental model of the new dilution refrigerator. M_1 and M_2 are the upper and lower mixing chambers. The connecting tube is coiled to reduce the thermal contact between M_1 and M_2 through the tube wall.

separation curve at the temperatures T_1 and T_2 of the two mixing chambers. T_1 is lower than T_2, and X_1 is therefore smaller than X_2 (*fig. 17*). The quantity $\dot{n}_{32}$ is thus positive, so that M_2 is in fact cooled. However, the phase-separation curve is almost vertical at $T < 0.1$ K. This means that $\dot{n}_{32}$ is almost zero, so that M_2 is not cooled beyond about 0.1 K. From that point onwards, M_2 acts as a heat exchanger, which brings the ^{3}He flowing from the pipe P_1 to a temperature of 0.1 K. The external circulation $\dot{n}_3$ is now fully available to contribute towards the cooling capacity in M_1.

Results

Although the new refrigerator, an example of which is shown in *fig. 18*, is still in an experimental stage, the following very provisional results indicate that it works well. The results were obtained with a model in which the superleak S_1 (see fig. 13) had not yet been connected through the heat exchanger W'' to the lower part of P_2 or with M_2, though it did have a heat exchange with the vaporization chamber V. This implies a fairly large heat leakage from M_1.

When we set the conventional part of the arrangement into operation we find, as with a conventional dilution refrigerator, that the inlet pressure of the external ^{3}He pump is about 100 Pa (1 mm Hg), corresponding to a circulation rate $\dot{n}_3$ of 1 mmol/s. In a conventional refrigerator the inlet pressure then decreases to 2 Pa, and $\dot{n}_3$ then falls to 10 μmol/s when the temperature of the mixing chamber goes below 0.1 K. In the new refrigerator, however, the fountain pump is brought into operation when the temperature T_2 of the lower mixing chamber M_2 reaches about 0.35 K. At that moment the temperature T_1 of the upper mixing chamber is about 0.7 K. After the fountain pump has started up, T_2 first rises to 0.45 K, while T_1 falls rapidly. When T_1 has reached about 0.45 K, both T_1 and T_2 decrease further. In this way a temperature T_1 of 20 mK can be reached about 10 minutes after the fountain pump has come into operation. The lower mixing chamber takes a longer time (half an hour or an hour) to reach the temperature of 0.1 K. In spite of the heat leakage, the lowest temperature T_1 we have reached is about 10 mK.

The cooling capacity can be calculated approximately from the rate at which T_1 decreases, since the specific heat of the helium present in M_1 is known. A calculation along these lines, at a given fountain-pump speed, gives the value quoted earlier for the cooling capacity: 1 mW at a temperature T_1 of 100 mK.

Summary. Since the sixties continuous cooling in the millikelvin range has become possible with 'dilution refrigerators', in which the heat of mixing is extracted from liquid helium present in two phases in a 'mixing chamber': a dilute solution of ^{3}He in superfluid ^{4}He, with a mixture rich in ^{3}He above it. The ^{3}He in the dilute solution is extracted through an upper vaporization chamber (still) at a higher temperature, and then returned to the mixing chamber after compression and recondensation. In refrigerators of this type the heat exchangers required between the upward and downward streams of ^{3}He are the main obstacle to the cooling process below 0.1 K. It is also difficult to circulate the ^{3}He rapidly and to avoid unwanted convection in the upward stream. A new dilution refrigerator is discussed in which these difficulties have been overcome. The circulating ^{4}He is injected into a mixing chamber filled with ^{3}He; the resultant dilute solution falls into a lower chamber which forms part of a conventional dilution-refrigerator circuit. The heat exchange between the liquids in the pipe connecting the two chambers is almost ideal. The ^{4}He carries the ^{3}He along with it in restrictions in the pipe, thus enhancing the ^{3}He circulation. So far a minimum temperature of about 10 mK has been reached and a cooling capacity of 1 mW at 100 mK, i.e. an order of magnitude higher than that of a conventional dilution refrigerator.

Recent 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*
Mullard Research Laboratories, Redhill, Surrey, England *M*
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*
MBLE Laboratoire de Recherches, 2 avenue Van Becelaere, 1170 Brussels (Boitsfort), Belgium *B*
Philips Laboratories, 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A. (by contract with the North American Philips Corp.) *N*

Reprints of most of these publications will be available in the near future. Requests for reprints should be addressed to the respective laboratories (see the code letter) or to Philips Research Laboratories, Eindhoven, The Netherlands.

V. Belevitch & Y. Genin: New results in one-port synthesis.
Philips Res. Repts. **30**, 360-383, 1975 (No. 5). *B*

C. Belouet, M. Monnier & J. C. Verplanke: Autoradiography as a tool for studying iron segregation and related defects in KH_2PO_4 single crystals.
J. Crystal Growth **29**, 109-120, 1975 (No. 1). *L, E*

F. Berz: Comments on 'Measurements and interpretation of low frequency noise in FET's'.
IEEE Trans. **ED-22**, 293-294, 1975 (No. 5). *M*

J. van den Boomgaard: Eutectic-like and peritectic-like reactions in ternary systems at a constant pressure, Part I. Systems containing one liquid and two solid phases.
Philips Res. Repts. **30**, 291-308, 1975 (No. 5). *E*

D. J. Breed: A new model for the negative voltage instability in MOS devices.
Appl. Phys. Letters **26**, 116-118, 1975 (No. 3). *E*

J. C. Brice: Some thermodynamic aspects of the growth of strained crystals.
J. Crystal Growth **28**, 249-253, 1975 (No. 2). *M*

A. Broese van Groenou, N. Maan & J. D. B. Veldkamp: Scratching experiments on various ceramic materials.
Philips Res. Repts. **30**, 320-359, 1975 (No. 5). *E*

H. H. Brongersma: Composition and structure analysis of surfaces by ion scattering.
Silicates ind. **40**, 49-54, 1975 (No. 2). *E*

M. Brouha & K. H. J. Buschow: Magnetic properties of $LaCo_{5x}Ni_{5-5x}$.
J. Physics F **5**, 543-554, 1975 (No. 3). *E*

H. Dimigen & H. Lüthje: The influence of the target material on sputter etching processes.
Thin Solid Films **27**, 155-163, 1975 (No. 1). *H*

J. Donjon, M. Decaesteker, B. Monod & R. Petit: Résolution spatiale des relais optiques utilisant l'effet Pockels.
Acta Electronica **18**, 187-199, 1975 (No. 3). *L*

J. W. F. Dorleijn & A. R. Miedema: A quantitative investigation of the two current conduction in nickel alloys.
J. Physics F **5**, 487-496, 1975 (No. 3). *E*

H. Durand: The Titus system: its characteristics, its performance, its market.
Acta Electronica **18**, 163-170, 1975 (No. 3). *(Also in French.)* *L*

R. G. Gossink, H. A. M. Coenen, A. R. C. Engelfriet, M. L. Verheijke & J. C. Verplanke: Ultrapure SiO_2 and Al_2O_3 for the preparation of low-loss compound glasses.
Mat. Res. Bull. **10**, 35-40, 1975 (No. 1). *E*

K. H. Härdtl: Gas isostatic hot pressing without molds.
Amer. Ceramic Soc. Bull. **54**, 201-205 & 207, 1975 (No. 2). *A*

J.-P. Hazan: Les applications de Titus et de Phototitus au traitement d'images et de données.
Acta Electronica **18**, 201-219, 1975 (No. 3). *L*

B. Hill, U. J. Schmidt & H. J. Schmitt: Optical memories.
J. appl. Sci. Engng. A **1**, 39-48, 1975 (No. 1). *H*

M. J. G. van Hout, J. C. Verplanke & J. M. Robertson: Hydrothermal synthesis of single crystal thin films of magnetic garnets and their analysis.
Mat. Res. Bull. **10**, 125-131, 1975 (No. 2). *E*

W. H. de Jeu & Th. W. Lathouwers: Dielectric constants and molecular structure of nematic liquid crystals, II. Variation of the bridging group.
Z. Naturf. **30a**, 79-82, 1975 (No. 1). *E*

Y. Kamp & J. P. Thiran: Chebyshev approximation for two-dimensional nonrecursive digital filters.
IEEE Trans. **CAS-22**, 208-218, 1975 (No. 3). *B*

J. P. Klozenberg: A theoretical comparison of two methods of supporting a pyroelectric target in a thermal imaging system.
Infrared Phys. **15**, 87-93, 1975 (No. 2). *M*

H. Koeman: Discussion on optimum filtering in nuclear radiation spectrometers.
Nucl. Instr. Meth. **123**, 161-167, 1975 (No. 1). *E*

H. Koeman: Principle of operation and properties of a transversal digital filter.
Nucl. Instr. Meth. **123**, 169-180, 1975 (No. 1). *E*

H. Koeman: Practical performance of the transversal digital filter in conjunction with X-ray detector and preamplifier.
Nucl. Instr. Meth. **123**, 181-187, 1975 (No. 1). *E*

B. Kramer & A. Mircea: Determination of saturated electron velocity in GaAs.
Appl. Phys. Letters **26**, 623-625, 1975 (No. 11). *L*

P. Kramer, C. de Vries & L. J. van Ruyven (Philips Semiconductor Development Laboratory, Nijmegen): The influence of leakage current on concentration profile measurements.
J. Electrochem. Soc. **122**, 314-316, 1975 (No. 2).

G. M. Loiacono: The industrial growth and characterization of KD_2PO_4 and CsD_2AsO_4.
Acta Electronica **18**, 241-251, 1975 (No. 3). *N*

G. Marie: Projection d'images de télévision sur grand écran.
Acta Electronica **18**, 221-231, 1975 (No. 3). *L*

G. Marie, J. Donjon, B. Monod & M. Decaesteker: Contraste des images projetées à l'aide de relais optiques utilisant l'effet Pockels.
Acta Electronica **18**, 171-186, 1975 (No. 3). *L*

F. Meyer & A. Kroes: Surface states at the clean surfaces of cleaved Si(111) and GaAs(110).
Surface Sci. **47**, 124-131, 1975 (No. 1). *E*

K. H. Nicholas, H. E. Brockman & I. J. Stemp: Fabrication of submicron polysilicon lines by conventional techniques.
Appl. Phys. Letters **26**, 398-399, 1975 (No. 7). *M*

W. J. Oosterkamp & A. P. M. van 't Hof: Colour radiography and subtraction.
Proc. XIII Int. Congress of Radiology, Madrid 1973, Vol. 2, pp. 387-391; 1975. *E*

H. I. Ralph, G. Simpson & R. J. Elliott (University of Oxford): Central-cell corrections to the theory of ionized-impurity scattering of electrons in silicon.
Phys. Rev. B **11**, 2948-2956, 1975 (No. 8). *M*

G. W. Rathenau: Again: depletion of natural resources.
Mat. Sci. Engng. **17**, 1, 1975 (No. 1). *E*

H. Rau: Defect model for cubic high temperature cuprous sulfide (digenite).
Solid State Comm. **16**, 1041-1042, 1975 (No. 8). *A*

U. Rothgordt: Run-length coding method for black-and-white facsimile with a ternary code as an intermediate step.
Electronics Letters **11**, 101-102, 1975 (No. 5). *H*

P. Schagen: Electronic aids to night vision.
Electronics & Power **21**, 437-439, 1975 (17 April). *M*

G. H. A. M. van der Steen (Philips Glass Development Centre, Eindhoven): Introduction and removal of hydroxyl groups in vitreous silica, Part III. Some thermodynamic data on the reduction of vitreous silica.
Philips Res. Repts. **30**, 309-319, 1975 (No. 5).

A. L. N. Stevels & F. Pingault (Massiot-Philips, Ailly-Le-Haut-Clocher, France): $BaFCl:Eu^{2+}$, a new phosphor for X-ray-intensifying screens.
Philips Res. Repts. **30**, 277-290, 1975 (No. 5). *E*

W. Thust (Elektro Spezial, Bremen): Laser beam deflection in large scale display writing.
Acta Electronica **18**, 233-239, 1975 (No. 3).

J. J. Vrakking & F. Meyer: Measurement of ionization cross-sections and backscattering factors for use in quantitative AES.
Surface Sci. **47**, 50-56, 1975 (No. 1). *E*

J. W. ter Vrugt, L. W. J. Manders & W. L. Wanmaker (Philips Lighting Division, Eindhoven): Fluorescent lamp coatings with layers of different compositions.
Lighting Res. Technol. **7**, 23-29, 1975 (No. 1).

W. L. Wanmaker (Philips Lighting Division, Eindhoven), **J. F. van Gils** (Erasmus University, Rotterdam) **& J. W. ter Vrugt** (Philips Lighting Division, Eindhoven): New blue lamp for phototherapy of hyperbilirubinemia.
Lighting Res. Technol. **7**, 19-22, 1975 (No. 1).

H. J. Weegenaar (Centraal Beheer - CEA, Apeldoorn) **& D. K. Wielenga:** Towards a generalized command language for job control.
Command languages, editor C. Unger, North-Holland Publ. Co., Amsterdam 1975, pp. 191-206. *E*

H. Weiss: Nonredundant point distribution for coded aperture imaging with application to three-dimensional on-line X-ray information retrieving.
IEEE Trans. **C-24**, 391-394, 1975 (No. 4). *H*

K. Weiss: Kriechvorgänge in einkristallinem NaCl.
Z. phys. Chemie neue Folge **94**, 273-286, 1975 (No. 4-6). *E*

H. W. Werner: The use of secondary ion mass spectrometry in surface analysis.
Surface Sci. **47**, 301-323, 1975 (No. 1). *E*

J. M. Woodcock, J. M. Shannon & D. J. Clark: Electrical and cathodoluminescence measurements on ion implanted donor layers in GaAs.
Solid-State Electronics **18**, 267-275, 1975 (No. 3). *M*

PHILIPS TECHNICAL REVIEW

VOLUME 36, 1976, No. 5

Electro-optic deflection of a laser beam

U. J. Schmidt

Although it seemed at first that the laser would have many technical applications, they have been relatively few in number until a few years ago. At Philips Forschungslaboratorium Hamburg a group of scientists have designed and developed a practical method for deflecting a laser beam in a large number of fixed directions ('digital deflection'). This method offers a whole range of possible applications, from magneto-optic computer memories to data projection on a screen several square metres in area, with 10^6 picture elements. One of the prototypes of such a display is already in use elsewhere in Germany.

A new technique, known as digital light deflection, allows a beam of laser light to be deflected very rapidly. The deflection is controlled by electrical signals, without the need for any moving mechanical components. The method makes use of classical physical effects and can in principle be used to deflect any light beam, but owes its significance to the special properties of laser beams. After a certain amount of refinement, this new electro-optic technique will make the laser available for application in many different fields of technology. These could include telecommunication, mass memories [*], the manufacture of microfiches, and large-screen projection.

This article will be chiefly concerned with the deflection method developed at Philips Forschungslaboratorium Hamburg. As an example of an important application, a large-screen projector controlled by a computer will be described. This gives a picture several square metres in area and consisting of 1024×1024 (i.e. 2^{20}) picture elements. Before going on to discuss this application, we shall first consider the physical basis of our deflection method, as well as some of the technical aspects of the equipment and some of the details concerning the design of the 20-stage deflector that forms the heart of the large-screen projector [1].

Dr U. J. Schmidt, now with the Philips Data Systems Division at Eiserfeld, West Germany, was formerly with Philips GmbH Forschungslaboratorium Hamburg, Hamburg, West Germany.

The physical basis of digital light deflection; the principle of a 20-stage deflector

Fig. 1a gives the starting point for our method of digital deflection [2]: the double refraction in a prism, preferably ground from a calcite crystal, in such a way that the refracting edge of the prism is parallel to the optical axis. With this arrangement a well known experiment can be performed: a beam of light incident on the prism will be seen to separate into *two* beams, the 'extraordinary' beam (*1*) and the 'ordinary' beam (*2*). The extraordinary beam is linearly polarized parallel to the optic axis, the ordinary one perpendicular to it. If the incident beam is itself already linearly polar-

[*] An article on this will appear in the next volume of Philips Technical Review. *(Ed.)*

[1] Further descriptions of this application can be found in:
U. J. Schmidt, Present state of the digital laser beam deflection technique for alphanumeric and graphic displays, in: E. Camatini (ed.), Progress in electro-optics, Plenum Press, New York 1975, pp. 161-179;
W. Thust, Grossbilddarstellung mit Laser, Umschau **75**, 374-375, 1975;
W. Thust, Laser beam deflection in large scale display writing, Acta Electronica **18**, 233-239, 1975;
E. Schröder, Digitale Laserstrahl-Richtungsmodulation und ihre Anwendungen, Feinwerktechnik + Micronic **78**, 64-68, 1974;
E. Schröder, Digital electro-optic light deflection, 4. Int. Kongreß für Reprographie und Information 1975, Zusammenfassung der Vorträge, pp. 40-42.

[2] U. J. Schmidt, in: D. K. Pollock, C. J. Koester and J. T. Tippett (ed.), Optical processing of information, Spartan Books, Baltimore 1963, p. 98.

ized parallel to the optic axis, the light only propagates in direction *1*; if the beam is polarized perpendicular to the optic axis, it only propagates in direction *2*. This effect we now put to use. It explains why the deflection method that we describe here is known as 'digital'. By orienting the polarization plane so that either the one or the other of the two orthogonal polarization states arises, the incident light beam can be deflected in either the first direction or the other.

based on various physical effects, but as yet the only practical type that will give speeds into the μs range or faster is the well known Kerr cell. The essential components of a Kerr cell (*fig. 2*) are two planar electrodes, of length *l*, placed opposite each other and with an electro-optic liquid between them. When a voltage is applied across the electrodes, the liquid becomes birefringent. The state of polarization of light passing through the cell then undergoes a change. The phase

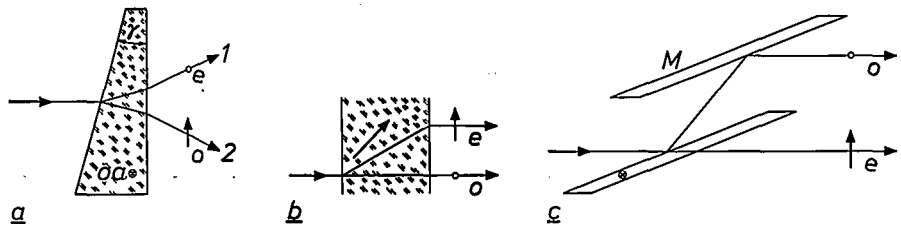


Fig. 1. *a*) When a beam of light is incident on a prism of birefringent material (refracting angle γ) whose optic axis (*oa*) is parallel to the refracting edge, the beam separates into a beam linearly polarized parallel to the optic axis (beam *1*, the 'extraordinary ray'), and a second beam, polarized perpendicular to the optic axis (beam *2*, the 'ordinary ray'). If the incident beam is itself linearly polarized in one of these two directions, then only one beam is transmitted. By rotating the plane of polarization of the incident beam through 90°, the transmitted beam can be made to switch from direction *1* to direction *2*, and vice versa ('digital deflection'). *b*, *c*) A separation into two parallel beams can be obtained by using a birefringent crystal with two parallel surfaces and the optic axis in the direction of the arrow, or by using a birefringent plate and an ordinary mirror *M*, with the angle of incidence such that the ordinary ray (*o*) is reflected and the extraordinary ray (*e*) is just not reflected. In cases (*a*) and (*c*) the elements must be located in a medium of suitable refractive index.

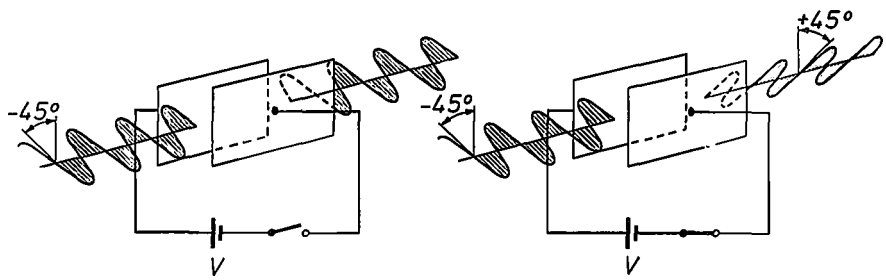


Fig. 2. Principle of the construction of a Kerr cell. Two parallel electrodes are located in an electro-optically active liquid. When a voltage *V* is applied, the liquid becomes birefringent. A phase difference dependent on the applied voltage then appears between the component of the emergent beam polarized parallel to the electric field and the component polarized perpendicular to the field. If the incident light is linearly polarized with the plane of polarization in the state shown, and the voltage is selected to give a phase difference of 180°, then the emergent beam is also linearly polarized, in a plane perpendicular to the first one.

The principle just described is not the only one that can be applied in the digital deflection of a laser beam. Two other solutions have been quoted in the literature [3] [4]. These are indicated schematically in fig. 1*b* and 1*c*; in both cases a birefringent plate is substituted for the birefringent prism. The two possible polarization states are then used to give a beam displacement by birefringent refraction or a deflection as a result of birefringent reflection.

The polarization plane can be made to rotate in a number of ways. Because of the high-speed operation required in many applications, it is necessary to use electro-optic polarization modulators with no moving parts. Polarization modulators can in principle be

difference Φ between the amplitude component of this beam parallel to the applied field and the component perpendicular to it is given by Kerr's law:

$$\Phi = Bl(V/d)^2, \qquad (1)$$

where B is Kerr's constant and d is the spacing between the electrodes. With such a device an incident linearly polarized beam can be converted into a linearly polarized emergent beam whose plane of polarization makes an angle of 90° with that of the incident beam. This is done by setting these directions of polarization at an angle of 45° to the direction of the electric field and also applying a voltage such that the phase difference

between the two components of the emergent beam is exactly equal to 180° ('½λ shift'). It follows from (1) that this voltage must be equal to $d(\pi/Bl)^{\frac{1}{2}}$.

The combination of a polarization switch and a prism, as shown in *fig. 3a*, is the basic element of our devices for deflecting a laser beam: we call the combination a deflector stage. Fig. 3*b* illustrates the vital step that took us from the simple experiment of fig. 1*a* to the concept of the digital deflector; the two directions in which the first stage deflects the light can each be doubled by adding a second stage, which splits the light by an angle twice as large as the first stage. To generalize, a series of N stages, with the splitting angle of each successive stage twice that of the one before, allows the beam to be directed towards the 2^N different addressable locations. *Fig. 4* shows an extension of the idea, in which a linear deflection is converted to one in two dimensions, for a six-stage deflector. The prisms in three of the stages are rotated by 90° about the axis

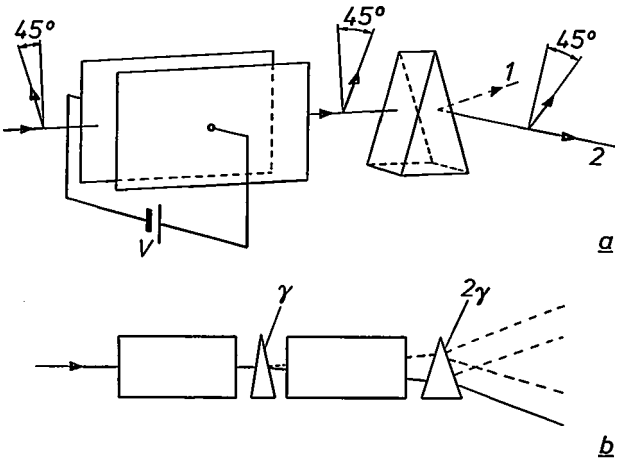


Fig. 3. *a*) A deflector stage consists of a combination of a Kerr cell (fig. 2) and a calcite prism (fig. 1*a*). The voltage V across the cell determines whether the emergent beam from the stage will have the direction *1* (from fig. 1) or the direction *2*. *b*) A simple two-stage deflector. The deflection obtained with the first stage can be doubled by the second if the refracting angle of the prism in the second stage is twice that of the prism in the first.

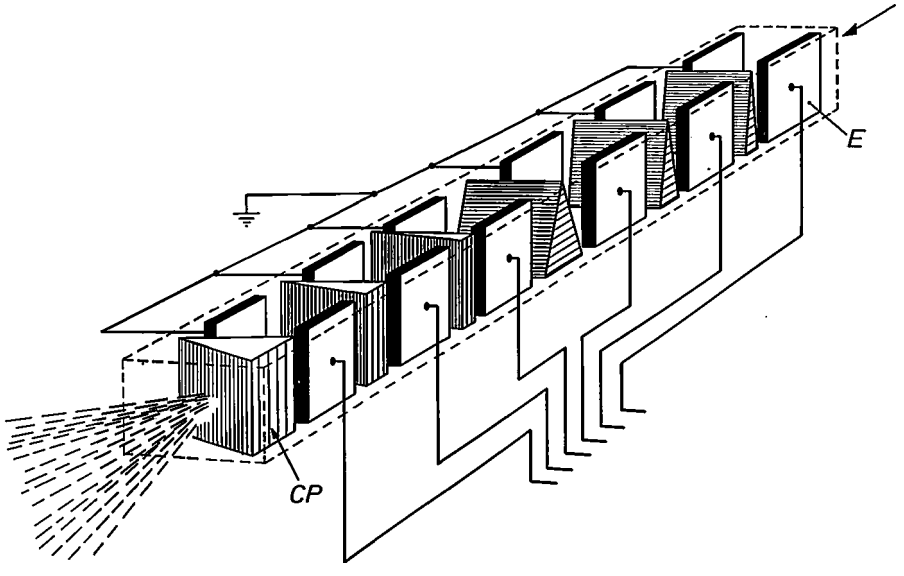


Fig. 4. A six-stage deflector (schematic) for deflecting a laser beam in two directions. The calcite prisms (*CP*) and the electrodes (*E*) of the Kerr cells are contained in a glass trough filled with nitrobenzene.

of the system; their refracting angles are the same as those of the other three prisms. The square scan pattern of 8×8 locations provided by this six-stage deflector can of course be extended in powers of two in the x- and y-directions independently by adding stages.

In recent years a number of experimental deflectors based on the principle of fig. 4 have been constructed. Deflectors constructed and tested so far have had a maximum of 20 stages; in the meantime it has been found that 22-stage deflectors are also a technical possibility. *Fig. 5* is an illustration of one of our 20-stage deflectors; more technical detail about this will be given later.

A third dimension could be added to the movement of the beam focus by complementing the x-y deflector with a third series of stages, which contain birefringent lenses in place of birefringent prisms [2]. The optic axes of these lens elements would then have to be parallel to those of the prisms or perpendicular to them. The operation of the stages with the lenses allows the collimated laser beam to be changed to a convergent or a divergent beam. Just as before, 2^N image locations in the direction of the axis are obtained, provided that there are N of these focusing stages.

[3] W. Kulcke, T. J. Harris, K. Kosanke and E. Max, IBM J. Res. Devel. 8, 64, 1964.
[4] T. J. Nelson, Digital light deflection, Bell Syst. tech. J. 43, 821-845, 1964.

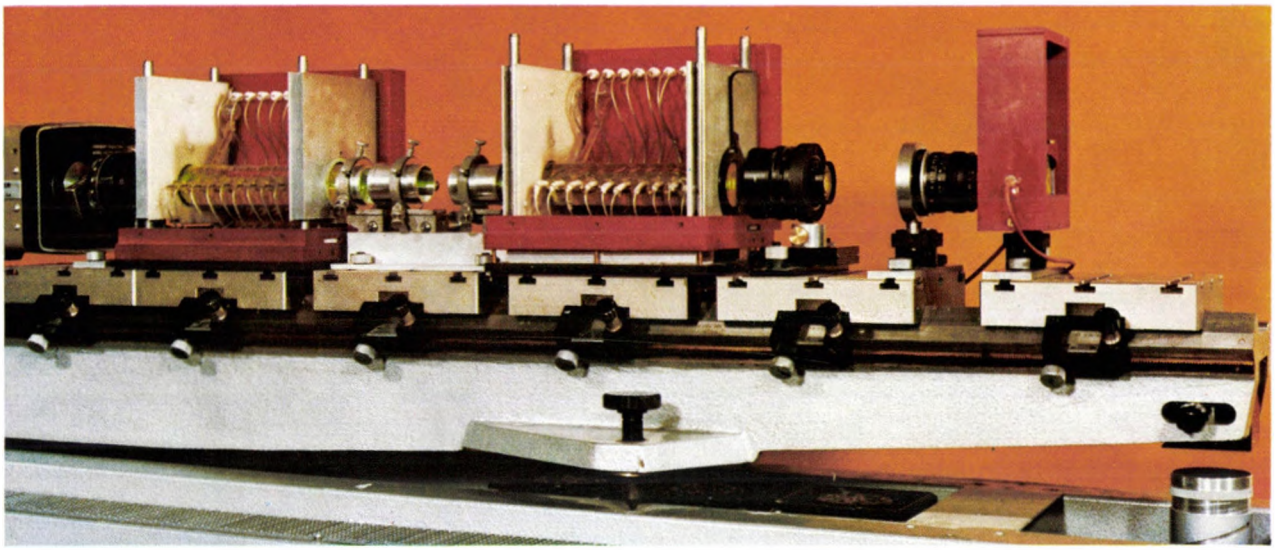


Fig. 5. Photograph of a 20-stage deflector. The deflector consists of two sections of ten stages, each constructed in accordance with the principle shown in fig. 4. The thin leads carry the supply to the Kerr cells. Part of the laser can be seen at the far left. The unit at the far right is a dynamic polarization filter — i.e. a combination of a linear polarizer and a Kerr cell — for suppressing cross-talk.

Features and problems

The deflection method illustrated in fig. 3 and fig. 4 possesses a number of attractive features. First of all there is the fact that the pattern of discrete beam positions is defined by the refracting angle of the birefringent prisms. A deviation from the required voltage will therefore not affect the position of the deflected laser beam, but only introduce a certain amount of cross-talk. This arises because an incorrect voltage gives the laser beam a polarization lying between the two orthogonal stages of polarization, so that two beams emerge from the deflector stage: one is the signal beam, the other the cross-talk beam, which usually has only a low intensity. An expression for calculating the cross-talk and a method for reducing it will be discussed later.

The second advantage is that each of the discrete beam positions is characterized by an unambiguous address; for an N-stage deflector this consists of N binary digits. Each digit describes the state of one of the Kerr cells. This means not only that the control logic for the laser beam can be very simple, but also that the electronic circuits of the deflection system can be directly connected to a digital computer.

If each Kerr cell of a deflector stage is energized by its own electronic voltage generator — this is the third important advantage — then the switching time is the same whether one Kerr cell or more are switched simultaneously. If this condition is satisfied, the deflector has random-access characteristics. Whatever the previous beam position, the switching time is always the same.

In the optical components, the use of birefringent prisms in the form of single elements limits the number of surfaces per stage in the light path to the absolute minimum of two. The reflection losses in our deflectors are only about 0.2% per stage. The absorption in calcite is low.

Compared with other possible forms of polarization switches the Kerr cell offers many advantages:

— Provided the temperature is the same everywhere, a liquid forms a body that is ideally homogeneous.

— The production of schlieren effects as a result of temperature gradients can be effectively suppressed by circulating the active liquid.

— Because a liquid is isotropic large slope angles are permissible between the beam and the axis of the system.

— A liquid does not give a piezoelectric effect.

— Both real and imaginary parts of the permittivity are small for many liquids, including nitrobenzene.

— The absorption of many liquids is small in the visible spectrum — this is also true for calcite.

The greatest problem with liquids is the small value of Kerr's constant. Nitrobenzene, for example, which is one of the most sensitive liquids available, requires an electric field-strength of 35 kV/cm. This is much higher than the 100 V/cm that is sufficient for modulation by the applied field in some modern electro-optic crystals. The difference is however partly cancelled out by the fact that the permittivity of these crystals is at least two orders of magnitude greater than those of the liquids. This means that the electric field energy required in a polarization switch with a liquid

is not much greater than in a switch with a crystal. When the various positive features listed above for the deflection with liquid electro-optic modulators are set against the characteristics of the solid-state deflectors, the general conclusion would appear to be that the liquid type is the more easily produced in practice. (This conclusion might no longer be valid if other requirements were set for the speed, resolution, etc.) The experimental results obtained so far seem to confirm these views: operating versions of the solid-state deflector with up to 14 stages have been described [5] [6], whereas 22-stage liquid-type deflectors have been made [7].

So far the discussion has been limited to the modulation of the direction of propagation of the collimated beam. In most applications it is necessary to convert the pattern of directions into a pattern of positions. This can be done by placing a lens behind the deflector. The beam will then be focused at the focal plane of the lens. With a laser beam of zero (spatial) order the intensity distribution at the picture element is Gaussian. The full width of this distribution, taken between $1/e^2$ intensity points, is equal to $4\lambda f/\pi w$, where f is the focal length of the lens, λ the wavelength and w the diameter of the beam at its smallest cross-section (the 'waist').

In the next section we shall look at a number of the many technical problems encountered in the design of a liquid-type deflector.

Technical design

The technical problems encountered in the construction of a digital deflector using single calcite crystals and Kerr cells filled with nitrobenzene can be roughly divided into two groups: those related to the properties of the electro-optic liquid and those related to the characteristics of the electronic circuits. There is also the requirement for small mechanical tolerances in the alignment of the prisms and the spacing of the electrodes. This is necessary because of the need for accurate location of the beam and for low cross-talk. We shall look first at the requirements for close tolerances in the electrode spacing.

If the relative 'cross-talk intensity' is defined as $\Delta I/I_0$, then the following expression can be derived for the permissible relative tolerance in the electrode separation:

$$\Delta d/d_0 = \sqrt{m\pi/\{m\pi \pm \arccos(1 + \Delta I/I_0)\}} - 1. \quad (2)$$

In this expression $m\pi$ is the 'prerotation' of the plane of polarization — this will be discussed later; m is an integer. *Table I* gives a few numerical values for the tolerance in the separation as a function of m.

Table I. The maximum permissible deviation ($\Delta d/d_0$) from the ideal spacing of the electrodes of a Kerr cell for cross-talk of 1% ($\Delta I/I_0 = 0.01$) as a function of the prerotation of the polarization plane ($m \times 90°$) produced by a fixed voltage.

m	$+\Delta d/d_0$	$-\Delta d/d_0$
1	0.028	0.033
2	0.016	0.017
3	0.010	0.011

In the adjustment of the birefringent prisms, it is found that the greatest accuracy is required in setting the angular orientation about the axis of the system, particularly for the prism with the largest splitting angle. Considering the prism of the Nth stage we find that the maximum permissible angular deviation θ is determined by the relation

$$\theta \leqslant 0.2/2^{N-1}. \quad (3)$$

If the displacement in a row of beam positions is not allowed to exceed 20% of an elementary deflection angle, then this means that θ in a 20-stage deflector should not be greater than 1.5′.

The magnitude of the voltage pulses necessary for the operation of the Kerr cells — up to 8.5 kV in a 20-stage deflector — necessitates a liquid of low conductivity, and hence of the lowest possible impurity content. If the resistivity of the liquid is lower than about 5×10^{10} Ωcm, the current in each Kerr cell has a value greater than 1 μA, and schlieren effects will therefore arise. These completely upset the optical homogeneity of the cell, which is the principal advantage of the liquid cell over the solid-state type.

The necessity for high purity is in itself a difficult requirement, and indeed we had to develop a special procedure for purifying commercially available nitrobenzene [8]. But the problem is intensified on two other counts: nitrobenzene of the desired conductivity is first of all highly hygroscopic and secondly it is a very good solvent — because of the high dipole moment of the molecules. For these reasons the cells in our deflectors are of glass, a chemically clean substance, which will cause hardly any 'doping' of the nitrobenzene. In addition the cells are hermetically sealed. The production of the glass vessels necessitated the development of a special high-precision construction technique. *Fig. 6* shows such a cell, for a 10-stage deflector. In the next subsection we shall look more closely at the problems of the electrical circuits.

[5] G. Hepner, IEEE J. QE-8, 169, 1972.
[6] P. Waterworth and D. C. J. Reid, Proc. First European Electro-Optics Markets and Technology Conf., Geneva 1972, p. 248.
[7] E. Schröder, Proc. Seminar Elektro-Optik, Munich 1972, p. 144.
[8] U. Krüger, R. Pepperl and U. J. Schmidt, Proc. IEEE 61, 992, 1973.

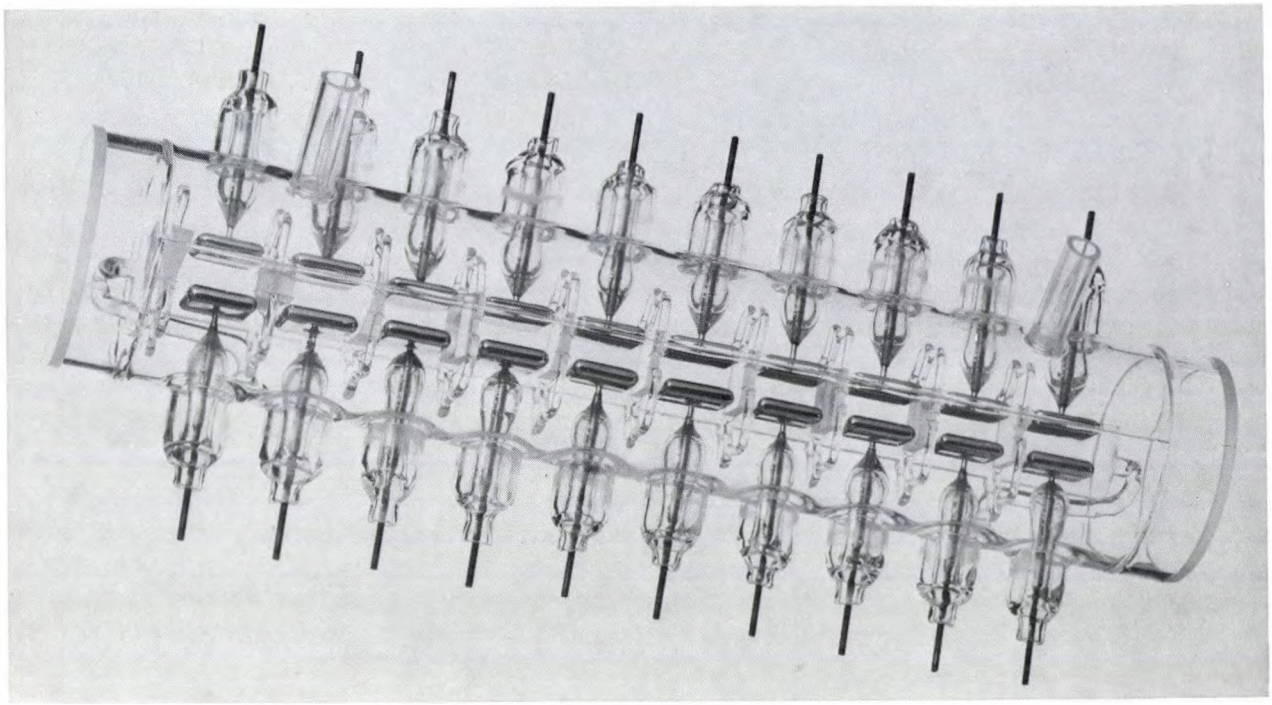


Fig. 6. Glass trough with ten sets of electrodes and ten calcite prisms; when the trough is filled with nitrobenzene ten deflector stages are produced.

To obtain some idea of the magnitude of the effect of inhomogeneities in the material that the laser beam passes through, we have performed model calculations on the assumption that the diameter of the inhomogeneities is small compared with the beam diameter and that the distribution of the phase changes caused by these inhomogeneities is Gaussian [9]. The intensity distribution in the image plane can then be expressed by a power series in σ, the standard deviation of the Gaussian curve for the phase variations in the wavefront of the beam. The first term of this series describes the original beam (the zero-order mode), the remaining terms describe the scattered light. *Fig. 7* gives a graphical representation of the first term as a function of σ after the beam has passed through one stage of a 20-stage deflector. If for example scattering losses of 5% are permissible, the r.m.s. phase change must remain below $\lambda/110$. This is a value that is very difficult to achieve with synthetic electro-optic crystals, as proposed for solid-state deflectors. The graph shows that even for much higher scattering losses, e.g. 50%, the desired crystal perfection should correspond to $\lambda/30$, which is still an exacting requirement.

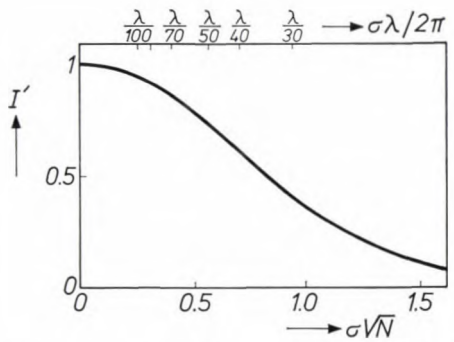


Fig. 7. Light loss due to scattering in a detector with N stages. The relative intensity I' of the beam (i.e. of the unscattered light) is plotted against the quantity $\sigma\sqrt{N}$, where σ is the r.m.s. value of the phase variations in the wavefront of the beam that are introduced in a single stage of a 20-stage deflector. With the upper scale the curve shows how the relative intensity of a beam that has passed through a single stage depends on σ.

The electrical circuits

The design parameters for a 20-stage deflector make a voltage of 20 kV necessary for the last Kerr cells of the deflector. To allow operation at a much lower control voltage use has been made of the nonlinear (square-law) variation of the Kerr effect; see *fig. 8*. If a bias voltage is applied such that the polarization plane of the light passing through the cell is given a 'prerotation' of an integral number of times 90° — let this integer be m — then the control voltage of the cell will be reduced by a factor of $(m+1)^{\frac{1}{2}}$. For $m = 1$, for example, this voltage is 0.4 of the bias voltage.

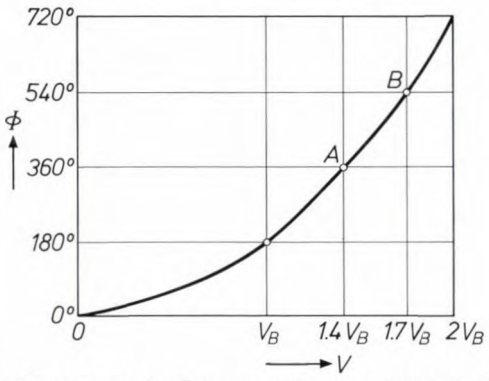


Fig. 8. The phase angle Φ between the components leaving the Kerr cell and polarized perpendicular and parallel to the applied electric field depends on the square of the applied voltage V. A phase angle of 180°, necessary for a 90° 'rotation' of the plane of polarization (see figs. 2 and 3), can therefore be achieved with a much smaller control voltage if a bias voltage is applied.

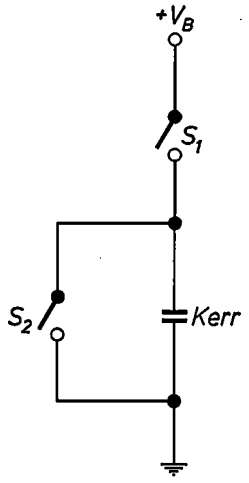


Fig. 9. Schematic diagram for the control of a Kerr cell with high voltage, without bias voltage (see fig. 8). The switches S must be capable of switching a voltage of several kV with speeds in the region of hundreds of kHz (see fig. 10).

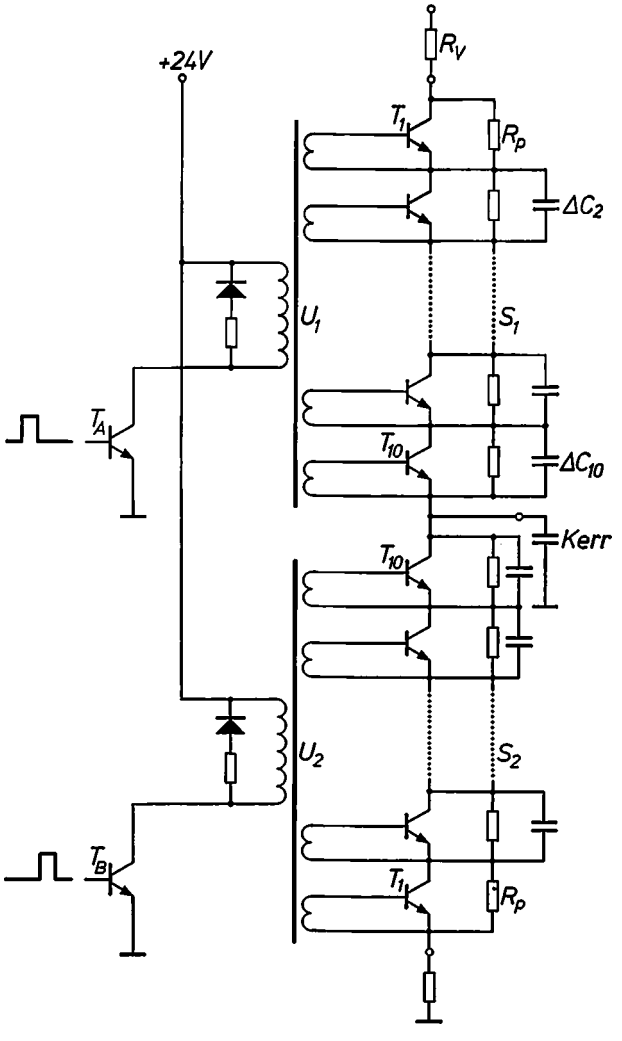


Fig. 10. Schematic diagram of the high-voltage supply for the control of the Kerr cells. The series of transistors T_1-T_{10} in the upper half of the figure performs the functions of the switch S_1 in fig. 9, and the series T_{10}-T_1 in the lower half that of S_2. The resistors R_p ensure a uniform voltage distribution when a group of transistors conducts, and the capacitors ΔC_i when the group does not conduct. U_1, U_2 pulse transformers. T_A, T_B transistors for control of the two switches. R_V bias resistor to limit the (dis)charge current.

In the design of the high-voltage supplies for the Kerr cells of a deflector use is made of the fact that only two voltage levels are necessary and that Kerr cells form a capacitive load. *Fig. 9* is a schematic diagram of the solution, which is essentially simple. Switch S_1 serves for charging the Kerr cell, switch S_2 for discharging it. If the high voltage V_B is not to be short-circuited during the beam switching, the combination of the two switches must operate so that one contact opens before the other closes. Such switches can be constructed in various ways. For the switching speeds mentioned earlier (in the range of hundreds of kHz) and for the switching voltages desired (up to 8.5 kV) the transistor circuits shown in *fig. 10* have been adopted. The transistors T_1-T_{10}, which together form a single switch, can each be put into the conducting state by a pulse from the corresponding secondary winding of the pulse transformer U_1 or U_2. The primary windings are supplied with pulses from low-voltage digital transistor circuits (T_A and T_B).

The chain of resistors and capacitors ensures that the distribution of the voltage across the transistors is uniform and constant. The resistances are made small enough for the current in them to be ten times as large as the leakage current in the non-conducting group of transistors. The capacitances ΔC are given values such that the effective capacitance to earth is the same for all the non-conducting transistors. This ensures that the capacitive current that flows when the other switch conducts produces the same voltage across all the transistors.

The switch (S_1 or S_2, fig. 9) that should conduct at a particular beam position starts to open again after each supply pulse (from T_A or T_B). The current i_R that starts to flow in the resistors will then tend to produce a voltage $\frac{1}{2}V_B$ across the Kerr cell, instead of V_B or 0. This undesired effect can be prevented by making the switch conduct sufficiently frequently; the intervals between the supply pulses are therefore made small in relation to CV_B/i_R, where C is the capacitance of the cell.

This problem of the correct charge distribution has further complications associated with it. In the arrangement shown in fig. 4 there is appreciable capacitive coupling between the adjacent Kerr cells. A change in the voltage across one cell therefore affects the voltage across the adjacent cells. To compensate for this effect, the switches associated with these adjacent cells should be conducting when there is such a change. For this reason, and for the one mentioned earlier, the switches S_1 and S_2 are opened at intervals corresponding to the

[9] U. J. Schmidt, E. Schröder and W. Thust, Appl. Optics 12, 460, 1973.

switching rate f of the laser beam. A further complication arises because the current for charging or discharging the Kerr cell is considerably higher than the current necessary to compensate for the capacitive coupling. The base signal for the transistors T_1 (fig. 10) is correspondingly adjusted, to ensure that no charge remains at the base after the transistors have stopped the current. Otherwise the transistor chain would continue to conduct — on account of the lifetime of the charge carriers at the base — while the other transistor chain was already starting to conduct for the next 'jump' of the laser beam.

Aberrations and their compensation

In this section a number of effects will be considered that would degrade the quality of the digital deflection if no countermeasures were taken: the cross-talk between channels, the influence of the operating temperature and the dispersion.

Cross-talk between channels

In switching the beam to a desired position the state of polarization should be purely linear and collinear with or perpendicular to the optic axis of the prisms in the deflector. These conditions will not be exactly satisfied in general. A number of effects have the result that a certain percentage of the light is polarized in a plane orthogonal to that of the signal beam. These include an incorrect control voltage, or an incorrect electrode spacing, field perturbations in the Kerr cells, incorrect alignment of the prisms, off-axis directions of the beam (apart from those in the 'horizontal' and 'vertical' plane) and inhomogeneities in the double refraction of the prisms. Assuming for simplicity that the degree of cross-talk is the same in each stage, we can easily calculate the cross-talk in the beam leaving the deflector. If the cross-talk arising in each stage is a fraction α of the original intensity of the beam, the intensity I_{sN} of the signal beam leaving the Nth stage is equal to $I_0(1-\alpha)^N$. We then find a power series in α for the cross-talk. The linear term in α describes light that is polarized orthogonally with respect to the signal beam; if $\alpha \ll 1$, this term will be by far the largest of the series. This allows the greater part of the cross-talk to be suppressed, by passing the deflected beam through a 'dynamic polarization filter' consisting of an extra Kerr cell and a fixed linear polarizer [10]. This Kerr cell is so arranged that it always keeps the signal beam polarized in a direction parallel to the direction of transmission of the polarizer. In this way the cross-talk light described by the odd terms of the series is suppressed. The total transmitted cross-talk C_0 when this countermeasure is not applied is equal to

Table II. The signal-to-background ratio (SBR) without and with the polarization filter.

Number of stages	Number of beam positions	Cross-talk per stage	SBR	
			Without filter	With filter
20	10^6	0.01	50	500
20	10^6	0.02	12.5	125
24	2×10^7	0.03	3.5	40

$$C_0 = \frac{I_0 - I_{sN}}{I_0} = 1-(1-\alpha)^N \approx N\alpha. \qquad (4a)$$

But if the polarization filter is used, there remains cross-talk

$$C_1 = 1-(1-\alpha)^N - N(1-\alpha)^{N-1}\alpha \approx \tfrac{1}{2}N(N-1)\,\alpha^2. \qquad (4b)$$

These expressions provide the respective signal-to-noise ratios directly:

$$SBR_0 = \frac{I_s}{I_0 C_0} = \frac{(1-\alpha)^N}{N\alpha} \approx 1/N\alpha \qquad (5a)$$

and

$$SBR_1 = \frac{I_s}{I_0 C_1} = \frac{(1-\alpha)^N}{1-(1-\alpha)^N - N\alpha\,(1-\alpha)^{N-1}} \approx 2/N^2\alpha^2. \qquad (5b)$$

The numerical values in *Table II* show the marked improvement when the polarization filter is used. This improvement is particularly remarkable when the cross-talk is less than 1% per stage.

The equations (4) and (5) are applicable when the cross-talk is at the same level in both states of the polarization switch. This is not always the case: for deflectors with nitrobenzene it is found that the lower

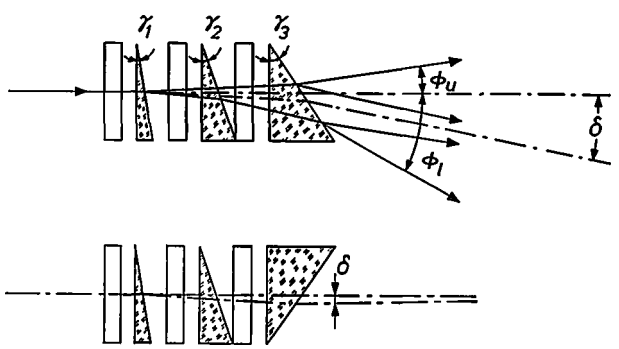


Fig. 11. Changes in the deflection pattern with temperature are mainly due to the temperature sensitivity of the nitrobenzene. These effects on the pattern can be almost completely eliminated by mounting the last calcite prism upside down. The angle δ is the change produced by temperature in the direction of the axis of the deflection pattern; ϕ_u angle between the upper deflection direction and the axis of the system; ϕ_l the same, but for the lower deflection direction.

state corresponds to an applied voltage of zero, with α then zero. In most of the experimental deflectors that we have made, however, a bias voltage is applied as explained in the previous section. Then α is not equal to zero and its magnitude corresponds to that in the higher state.

Compensation of the influence of the operating temperature

A deflector containing simple prisms, as shown in fig. 4, has the undesirable characteristic that a change in the operating temperature causes drift in the deflection pattern. This effect arises because in addition to the row of birefringent prisms there is also an associated row of prisms of nitrobenzene, whose refractive index varies fairly strongly with the temperature. A simple measure helps to reduce this effect very considerably [11]: instead of mounting all the prisms the 'right way' up, the last one is mounted 'upside down' (*fig. 11*). If we assume a linear relation between the angles of the prisms and their refraction — an approximation (of Snell's law) that is valid for small angles — the improvement obtained from this measure can be calculated directly. The change in direction $d\delta$ (per degree of temperature change) of the axis of the patterns of a system with no correction is proportional to the refracting angles γ_i of the prisms (which form a geometrical progression):

$$\frac{d\delta}{dT} \propto \sum_{n=1}^{N} \gamma_i = +(2^N - 1)\,\gamma_1. \qquad (6)$$

If the last prism is 'upside down', this equation takes the form

$$\frac{d\delta}{dT} \propto \sum_{n=1}^{N} \gamma_i = -\gamma_1.$$

The full equation reads:

$$\frac{1}{\eta_0}\frac{d\delta}{dT} = \frac{(2^N-1)}{(n_0-n_e)}\left\{\frac{1}{2}\frac{dn_e}{dT} + \frac{1}{2}\frac{dn_0}{dT} - \frac{dn_u}{dT}\right\}, \qquad (7)$$

where η_0 is the elementary splitting angle of the deflector, n_e and n_0 are the refractive indices of the birefringent material and n_u is the refractive index of the liquid surrounding the prisms. When calcite and nitrobenzene are used, and if the maximum permissible shift specified for a deflection pattern of 1024×1024 positions is set at no more than a tenth of the elementary splitting angle, a maximum permissible temperature change of ± 36 °C is obtained. If the last prism had not been inverted the value would have been ± 0.036 °C.

A more accurate approach shows that the permissible temperature variation will not be as large as the one just quoted, on account of the change that occurs in the refractive index of the prism material. This effect causes a change in the splitting angle with temperature in each prism. A detailed analysis shows that the sine function in Snell's law does have to be taken into account for the small changes that still remain. *Fig. 12* shows the numerical results of computer calculations that take these higher-order effects into account. It can be seen from the graph that the permissible change in temperature under the condition mentioned above is now ± 4 °C, a value still representing an improvement of about 100 times.

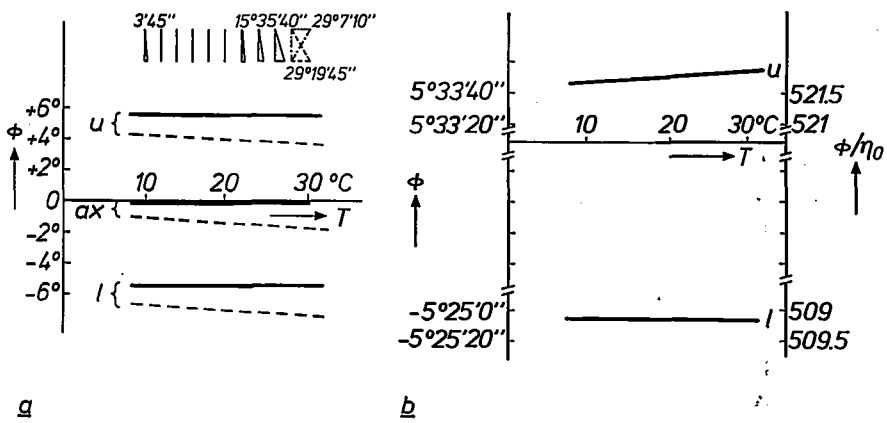


Fig. 12. *a*) Movement of the digitial deflection pattern with changing temperature in a 10-stage deflection system containing right-angled prisms. ϕ is the angle between the directions in question and the axis of the system. *u* upper beam position. *l* lower beam position. *ax* position of the axis of the deflection pattern. The dashed lines refer to the case in which the last prism is not upside down (see fig. 11), *b*) The same graph on an enlarged scale. η_0 is the elementary splitting angle.

[10] U. J. Schmidt, Physics Letters 12, 205, 1964.
[11] U. J. Schmidt and W. Thust, Opto-electronics 2, 29, 1970.

Dispersion effects

So far it has been tacitly assumed that the deflected beam is monochromatic. In some types of application it could well be desirable to use two or three colours. When light of more than one wavelength must be deflected by the same deflector, dispersion effects will affect the splitting angle of the individual deflector stages and also the position of the deflection pattern and the control voltage of the Kerr cell. The consequences of these three dispersion effects can be corrected for in a relatively simple way.

In the case where the deflector is controlled by a computer, the differences in position between picture elements associated with different colours can be reduced to zero by giving each position on the screen a corresponding address for each colour. If compensation to a fraction of the resolution angle is desired, extra deflection stages will be necessary for the associated small changes in position.

A second possible method would be to design an objective lens that had a dispersion exactly equal to that of the prism but of opposite sign. Calculations have shown that such a lens can be designed without too many complications [12].

The displacement of the axis of symmetry of the system with wavelength, the second dispersion effect mentioned above, can be almost completely compensated by inverting the last prism, as in eliminating the temperature sensitivity; the residual effect can be neglected.

The third dispersion effect, the wavelength dependence of the Kerr effect, can only be corrected by changing the control voltage when changing to another wavelength. Electronic circuits that can do this have been designed and constructed. *Fig. 13* shows a schematic circuit. The operation of the high-voltage switches S_i is in principle identical to that of the switches used in producing the fundamental control voltages, as described in the previous section.

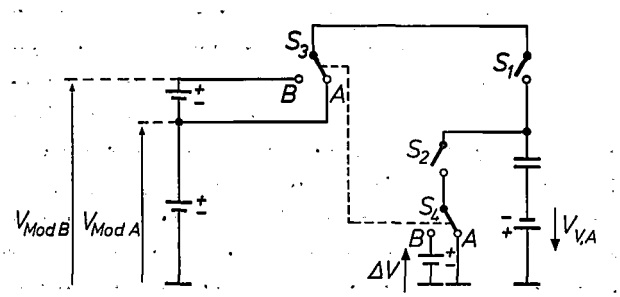


Fig. 13. For light beams of different wavelengths, the voltages V_{Mod} that make the plane of polarization rotate by 90° are also different. They can be obtained with the circuit shown. The high-voltage switches S are of the same type as those of fig. 10, and S_1 and S_2 also have the same function. $V_{V,A}$ bias for light of type A. Light of type B requires another bias, which is obtained by applying a correcting voltage ΔV via S_4.

Refraction aberrations

A further aberration should be taken into account when either the refracting angles of the birefringent prisms or the deflection angles become so large that the sine function rather than its argument should be used in Snell's law. The effect of this kind of aberration is an irregularity in the pattern of the deflection positions. A binary progression in the refracting angles of the birefringent prisms of a deflector, for example, would introduce

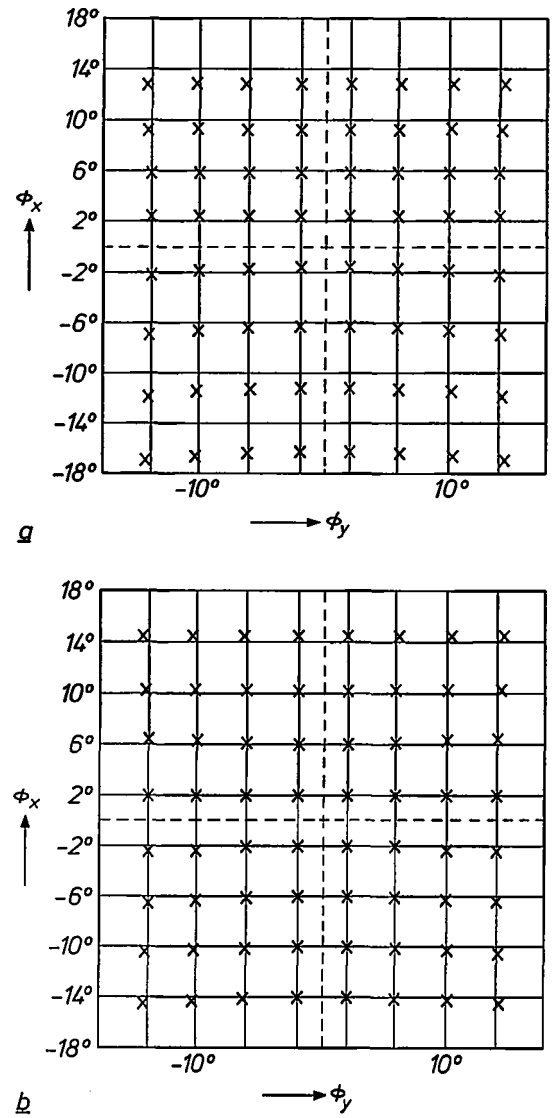


Fig. 14. *a*) If the refracting angle of the birefringent prisms or the deflection angles become so large that the sine function in Snell's law can no longer be approximated by its argument, 'pincushion' distortion appears in the deflection pattern. The example shown here is the deflection pattern of a Wollaston prism with an apex angle of 103°: 4×8 incident beams are doubled up to a field of 8×8 beams. *b*) The improvement gained by using two isosceles prisms with an apex angle of 128° (see fig. 15).

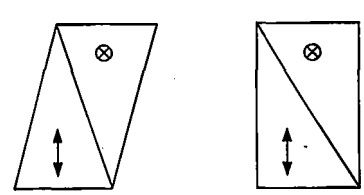


Fig. 15. Two methods of combining birefringent prisms to reduce distortion in the deflection pattern.

positional jumps between the corresponding subpatterns. A simple method for eliminating this aberration is to adjust the refracting angles of the prisms, going from those with the smallest angles to those with the largest, to bring the adjacent lines of the subpatterns concerned to distances corresponding to the desired elementary distance. This will not of course correct the non-linearity due to the varying angle of incidence, which causes not only a change in the distance in the x- and y-direction between adjacent subpatterns, but also in all intermediate directions, which in turn introduces distortions of the 'pincushion' type (*fig. 14a*). Compensation is not easy in this case, since there is no broad spectrum of birefringent materials available that is comparable with the available range for isotropic optical materials. The only type of correction available is the use of composite prisms made from a *number* of elements, as shown in *fig. 15* [13]. Fortunately this kind of correction is not found necessary in systems that have resolutions corresponding to a maximum of 2048 × 2048 positions and for any applications as yet considered [7].

splitting angles is of no great significance in most applications, since in general the beam is required to make jumps over larger distances less often than jumps over shorter distances.

The beam switching times obtained with these circuits range from 250 ns at voltages of 2.5 kV to 900 ns at 8.5 kV. *Fig. 16* shows oscillograms of the control-voltage pulses. In this experiment a d.c. bias corresponding to a prerotation of 90° for the plane of polarization of the light beam is applied across the Kerr cells. Owing to the square-law nature of Kerr's law, and the $\cos^2$ function for the intensity changes in the respective beam positions, the switching time of the light beam is about 7% smaller than that of the voltage.

Table III. Technical details of one of the experimental 20-stage deflectors.

Stage No.	Prism angle (γ)		Aperture		Electrode spacing (mm)	Voltage (for λ = 520.8 nm)	
	d stage	*h* stage	*d* (mm)	*h* (mm)		Bias (kV)	Control (kV)
1		6′	1.4	1.4	1.6	5.3	2.1
2		12′	1.4	1.4	1.6	5.3	2.1
3		24′	1.4	1.4	1.6	5.3	2.1
4		48′	1.4	1.4	1.6	5.3	2.1
5		1°36′	1.4	1.5	1.6	5.3	2.1
6	6′		1.4	1.6	1.6	5.3	2.1
7	12′		1.4	1.7	1.6	5.3	2.1
8	24′		1.4	1.8	1.6	5.3	2.1
9	48′		1.4	1.9	1.6	5.3	2.1
10	1°36′		1.5	2.0	1.6	5.3	2.1
11		1°36′	3.0	4.0	4.5	14	6.5
12		3°11′ 50″	3.0	4.1	4.5	14	6.5
13		6°22′ 20″	3.0	4.2	4.5	14	6.5
14	1°36′		3.1	4.6	4.5	14	6.5
15	3°11′ 50″		3.1	4.9	4.5	14	6.5
16	6°22′ 20″		3.3	5.3	4.5	14	6.5
17		12°35′ 15″	3.6	5.6	4.5	14	6.5
18	12°35′ 15″		4.0	6.4	4.5	14	6.5
19*		−25°11′ 30″	4.7	7.0	6.0	20	8.8
20*	−25°11′ 30″		5.5	8.6	6.0	20	8.8

* These two prisms are upside down to eliminate temperature effects.

Details and performance of a 20-stage deflector

In this section more details will be given of the 20-stage deflectors that we have made and tested [14], and which have been applied in the large-screen projection system that will be described in the next section. All the deflectors we have built (*Table III*) are designed for half-overlapping beam positions and they operate with electronic circuits like those shown in fig. 10. For control voltages up to 4.5 kV ten 2N3439 transistors are used in each switch, and for voltages up to 9 kV ten BU 205/208 transistors. The switching rates that can be attained with these transistors vary from 500 kHz at voltages less than 2.5 kV to 60 kHz at control voltages of 8.5 kV. The fact that the switching rates therefore become smaller for the deflector stages with the larger

The deflectors are usually used with the bias giving a prerotation of 90° in the polarization plane, as just mentioned. To prevent the occurrence of schlieren effects, we have kept the current below about 0.4 μA per stage: this corresponds to a resistivity of 3×10^{10} Ωcm. As a rule this value was not found to deteriorate as a function of the number of hours of operation. On the contrary, the resistivity usually improves with time in most cases. Measurements of the operation of the dynamic polarization filters — described in the pre-

[12] G. Kuttner, internal report, Rodenstock Optische Werke, 1973.

[13] U. J. Schmidt and W. Thust, Optik 32, 570, 1971.

[14] H. Meyer, D. Riekmann, K. P. Schmidt, U. J. Schmidt, M. Rahlff, E. Schröder and W. Thust, Appl. Optics 11, 1732, 1972.

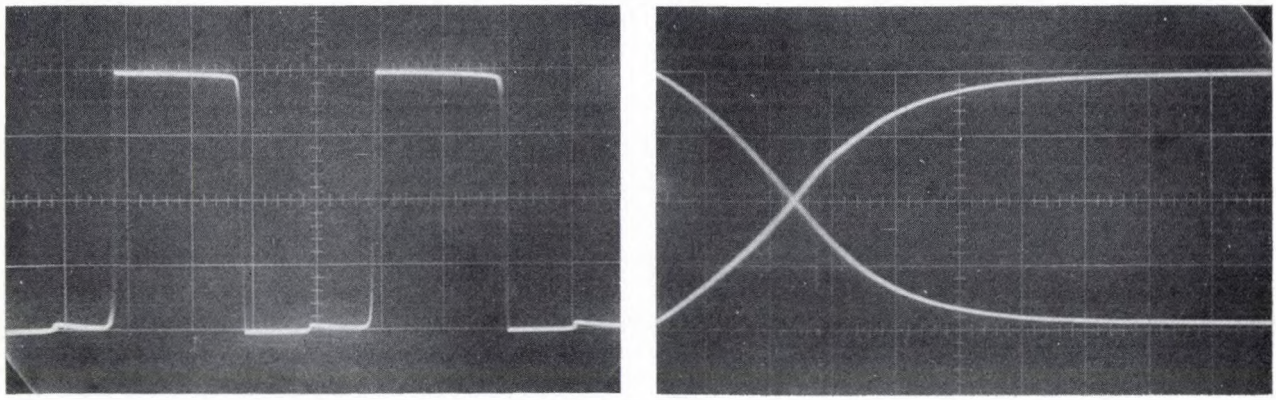


Fig. 16. Oscillograms of high-voltage pulses for a Kerr cell. The left-hand time scale is 10 μs per division, the right-hand one 0.2 μs per division. The right-hand picture shows that at a voltage of about 10 kV the switching time is about 1 μs. At lower voltages the switching time is shorter.

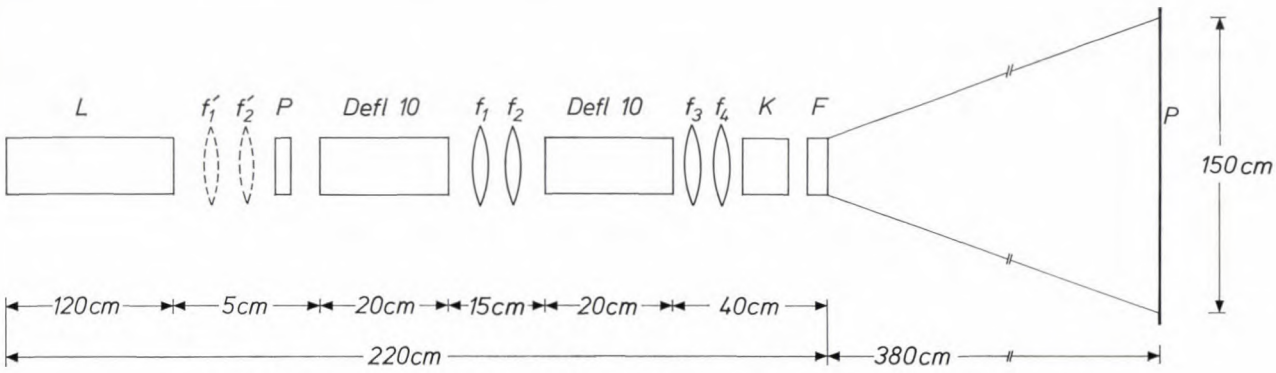


Fig. 17. The optical section of the 20-stage deflector shown in fig. 5, used in an experimental large-screen projector for alphanumeric data. *L* laser. *P* polarizer. *Defl 10* deflector with 10 stages. f_1 (40 mm) and f_2 (80 mm) telescope pair. f_3 (90 mm) and f_4 (35 mm) projection optics. *K* Kerr cell. *F* polarization filter. f_1' and f_2' matching lenses. *P* projection screen.

vious section — have indeed shown an appreciable reduction in the cross-talk between channels. If the power ratio of the strongest cross-talk beam and the signal beam is taken as a measure of the interference, we find that the operation of the dynamic polarization filter improves this ratio by a factor of about 20.

A large-screen projector for the presentation of alphanumeric and graphic data

We have constructed an experimental image-projection system, which is shown schematically in *fig. 17*. Its operation is simple in principle. To prevent flicker of the image, the laser beam is deflected at a field scan rate of about 60 Hz. The control signals can be derived from a digital memory of any type, in which each word of 20 bits represents a beam position. The data can be supplied by a process computer; its input data may be of any standard form. The projection rate is of course primarily determined by the switching rate of the beam, 500 kHz, for which the corresponding image has 8000 picture elements (= beam positions). It has been established experimentally that this capacity can be increased, at constant switching rate, by working with an irregular scanning pattern [15]. The random-access addressability of the deflector is then particularly advantageous. As an example, the effective scanning rate can be increased by a factor of two if for a first field only one of every two pairs of points is scanned, and the intermediate points form the following second field. In this way the capacity can be increased to 16 000 picture elements. A further increase can be obtained by using random-scanning methods like those used in image projection by cathode-ray tubes.

In spite of the high radiation intensity of laser beams, which makes them particularly suitable for deflection, their application in the projection of bright large images has not previously been very successful, because of the low efficiency of the laser process. The widely used gas lasers, for example, only have an efficiency of less than 0.01 %. Since pictures of television type — i.e. pictures scanned line by line — require an intensity of 1 W/m² for visibility in an illuminated room, about 20 kW of electrical power must be applied to each

[15] W. Thust, Int. Symp. Society for Information Display, San Francisco 1972, and Acta Electronica **18**, 233, 1975.

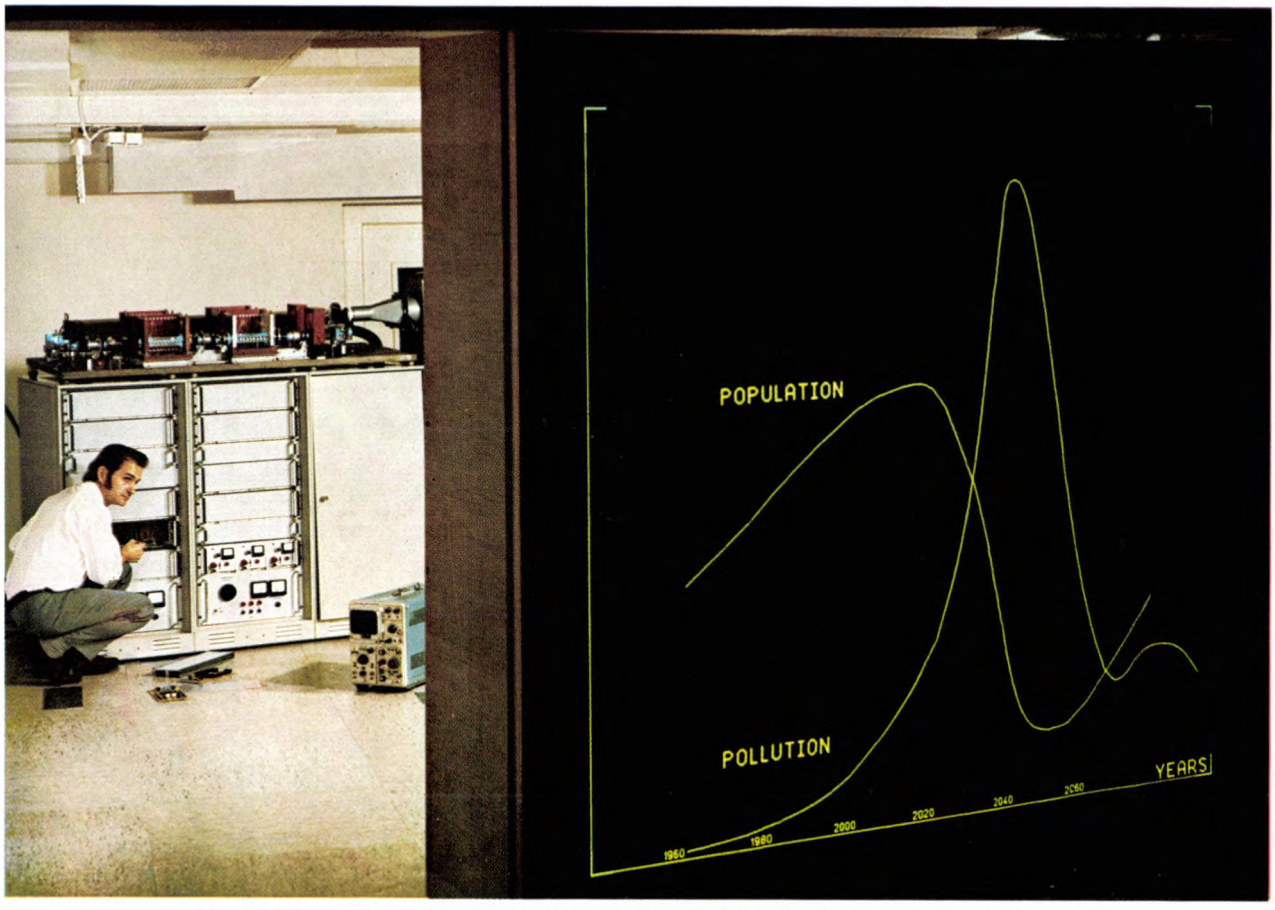


Fig. 18. The experimental system for large-screen projection with its screen (foreground). The screen measures 2×2 m^2 and is about 4 m from the deflection equipment.

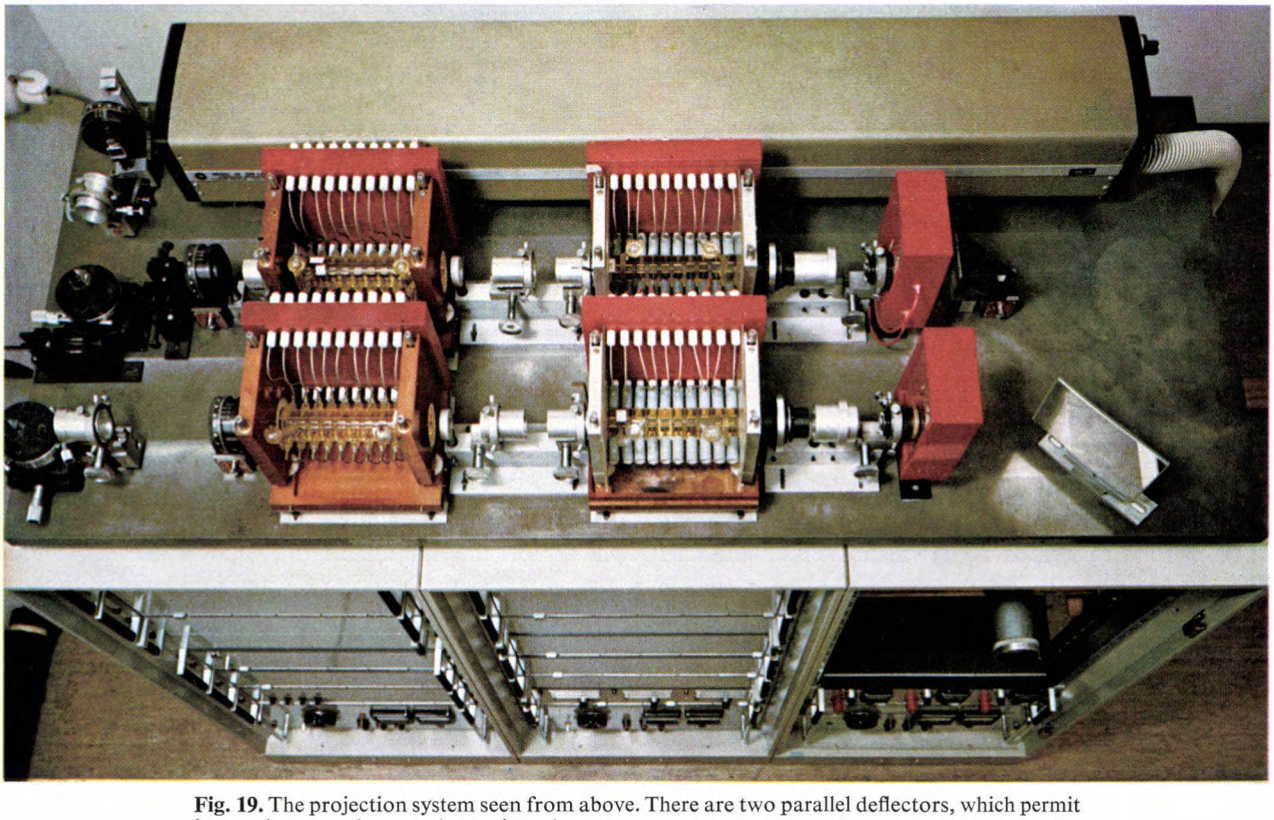


Fig. 19. The projection system seen from above. There are two parallel deflectors, which permit images in two colours to be projected.

square metre of the screen. (In this calculation it has been assumed that there is a light loss of 50% in the optical system.)

In the large-screen projector described here this quantity is much smaller, at least for producing images of graphs, because the equipment has random-access addressability. No beam suppression is therefore required, as would be the case in a line-scanning system. All the power that the laser can supply is used for the presentation of data. Compared with a line-scanning system there is an improvement equal to the ratio of the total image area to the part of the image area containing the information. This ratio can be as high as 50 for certain kinds of data. In general terms, 150 mW of laser power will give an illuminance of 1500 lux on a 2 m^2 screen. This is sufficient to allow the image to be observed in normal indoor illumination. *Fig. 18* and *fig. 19* are photographs of the experimental 20-stage deflector. A close-up of the screen is shown in *fig. 20*.

The problem of granulation

A problem that arises in all imaging systems with laser beams is that of granulation. This effect is present in all systems based on coherent optical imaging and containing scattering centres. It has been analysed and utilized extensively at many laboratories, including our own [16]. If no special precautions are taken, an observer looking at an image formed by laser light will notice that each picture element is composed of an irregular pattern of bright dots, separated by regions of zero intensity. The effect is due to interference between the elementary waves that are emitted by the scattering centres with a statistical spatial distribution, but coherent in time. Each movement of the observer's eye, or of the screen, which produces scattering, makes the pattern change.

Quite apart from the fact that the granulation is very annoying for the observer, it also degrades the effective resolution of imaging, since the 'median value' r of the dark regions increases with the distance of observation in accordance with the equation

$$r = L\lambda/p, \qquad (8)$$

where L is the distance between observer and screen, p the diameter of the pupil of the eye and λ the wavelength of the light. Above a particular distance the value of r exceeds the diameter of the picture element, which means that the element concerned is completely missing. Many attempts have been made to eliminate the granulation effect. This is a relatively easy matter for static images. Small movements of a suitably located scatterer, or of the actual scatterer that causes the granulation, will generally make the granulation disappear. This is because such movements cause a large number of statistically independent granulation patterns to move across one another, which 'averages out' the brighter and darker regions.

The theory indicates that the mean relative fluctuation in intensity $(\langle\langle(I - \langle I\rangle)^2\rangle\rangle)^{\frac{1}{2}}$ is related to the number of granulation patterns n by the equation [16]

$$\left\{\frac{\langle\langle(I - \langle I\rangle)^2\rangle\rangle}{\langle I\rangle^2}\right\}^{\frac{1}{2}} = n^{-\frac{1}{2}}. \qquad (9)$$

To give a pleasing laser image, about 100 patterns should be superimposed. This relatively large number is a particular problem with dynamic images, where the laser beam only remains in any position for a few microseconds, which means that the rate of change of the granulation patterns should be about 100 MHz.

In monochrome systems it is sufficient to coat the screen with a phosphor layer, so that the time coherence is destroyed by the conversion of the light [17]. This method is satisfactory up to the level at which the phosphor saturates. Often, however, the monochromaticity of the light is degraded to an extent that is undesirable for good reproduction.

A method for suppressing the granulation effect, which we have developed and applied at our laboratories, is to give a relative motion to two scatterers spaced by only a few microns [16]. This could be done by applying two scatterers to the bearing surfaces of a push-pull spiral-groove bearing [18]. Movement of the scattering centres over a correlation length gives a new granulation pattern. The two scatterers in our arrangement are located at an intermediate image plane (the common focal plane of the lenses f_3 and f_4 in fig. 17), so that a scattering area of only a few cm^2 is required. Since in practice the correlation lengths at this plane are only a fraction of a micron, velocities of about 10 m/s will be sufficient for dwell times of the order of 2 μs for the beam.

Finally, there is a third method that should be mentioned, in which no mechanical movement is required; see *fig. 21*. A scatterer, located at the plane *St* of the primary — almost microscopic — image, magnifies the angular aperture of the laser beam of spatial order zero such that it fills the aperture of the projection objective O_2. The wavefront of the beam at this objective is again that of a granulation pattern. Changing the phases of the individual granulation waves would change the granulation pattern of the imaging ray that arrives at the projection screen. This is done by placing an ultrasonic cell T at the objective O_2 and exciting the medium in the cell at a frequency whose wavelength corresponds approximately to the mean diameter of the granulation at the location of the ultrasonic wave. At a frequency of 500 kHz, for example, a complete cyclic process through a large number of granulation patterns

will take place within 2 μs. Various spatially independent ultrasonic waves will amplify this effect. Initial experiments have confirmed the effectiveness of this method. A disadvantage is the loss of intensity, by a factor of about two, because of the scattering process at the plane *St*. The method does however preserve the monochromaticity of the light, and is independent of the wavelength.

Appendix: The optimization of the geometry of the prism deflector

In designing a deflector system, it is of course desirable to try to optimize the design for parameters such as the resolution, the scanning rate, the power and the control voltage, the dimensions, the amount of material required, and so on. Some of these parameters are fixed as soon as a basic design has been drawn up, others are interrelated, and those still to be chosen cannot all be optimized at the same time. Just which ones should be optimized will also depend on the kind of application and the method of operation. A further consideration of the design approach shows that the important 'free' parameters that govern quantities such as control power or voltage are the sequence of the deflector stages, the geometry of the deflector and the diameter of the laser beam. In the following we shall determine the optimum geometry for minimum control voltage and for minimum control power, and we shall examine the extent to which the results are of importance for the design of a deflector [9] [19]. We shall begin by establishing the sequence of the deflector stages. Although a number of combinations could be considered, we shall only consider two here, shown in *fig. 22*, which will be sufficient to show how the performance of the deflector depends on the stage sequence. With the aid of fig. *22a*, we first determine the geometrical dimensions of the deflector if a laser beam of wavelength λ/n and diameter d (single mode TM_{00}) is to be deflected. The quantity d refers to the diagonal distance between the $1/e^2$ intensity points of the beam at its smallest cross-section. If the deflector is to be able to handle

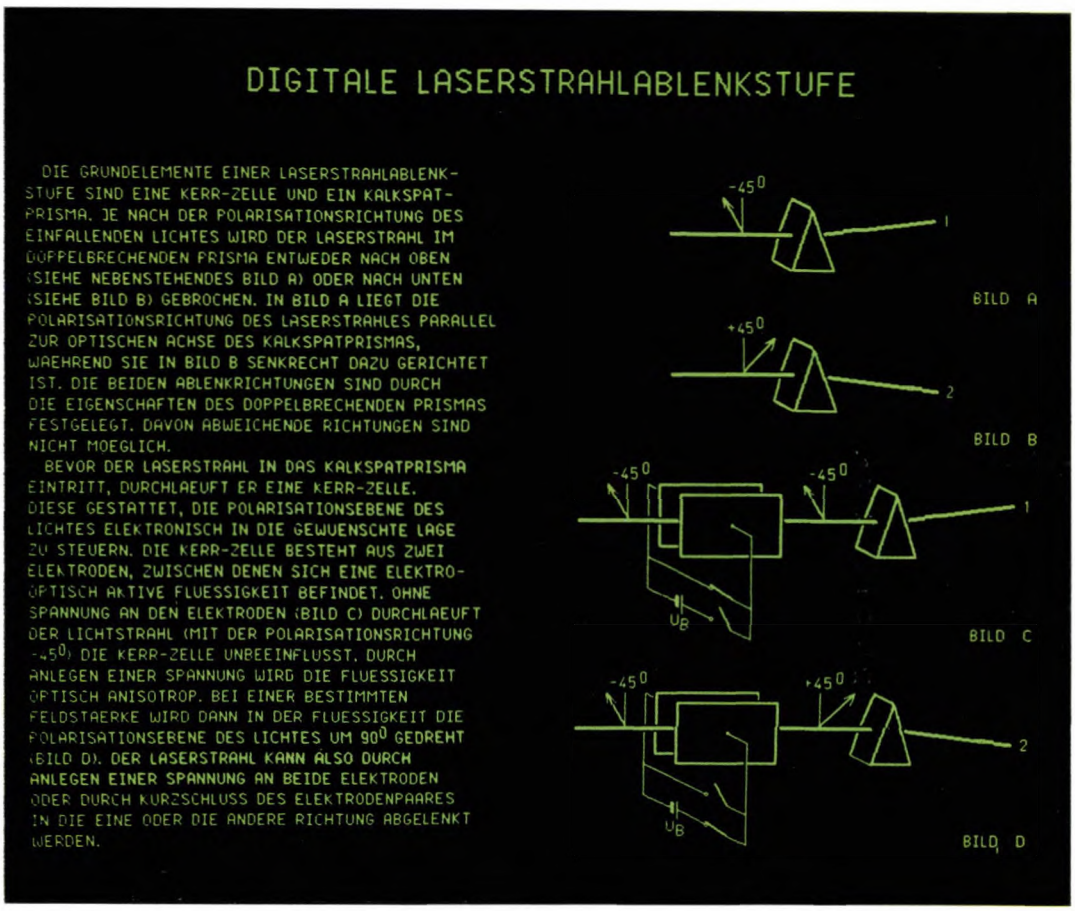


Fig. 20. Example of an image consisting of 40 000 points: scanning rate 12.5 Hz. The text explains the principle of operation.

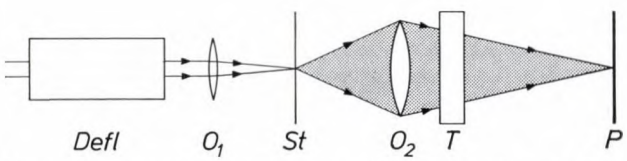


Fig. 21. Suppression of granulation with a scattering plate and an ultrasonic transducer. *Defl* deflector. O_1 lens forming an image on the scattering plate *St*. O_2 lens imaging *St* on the projection screen *P*. *T* ultrasonic cell.

[16] E. Schröder, Optics Comm. **3**, 68, 1971.
[17] L. G. Van Uitert, D. A. Pinnow and J. C. Williams, Appl. Optics **10**, 150, 1971.
[18] H. J. W. M. Volman, Philips tech. Rev. **35**, 11, 1975.
[19] A discussion of optimization in relation to the beam diameter is given in the article of note [14].

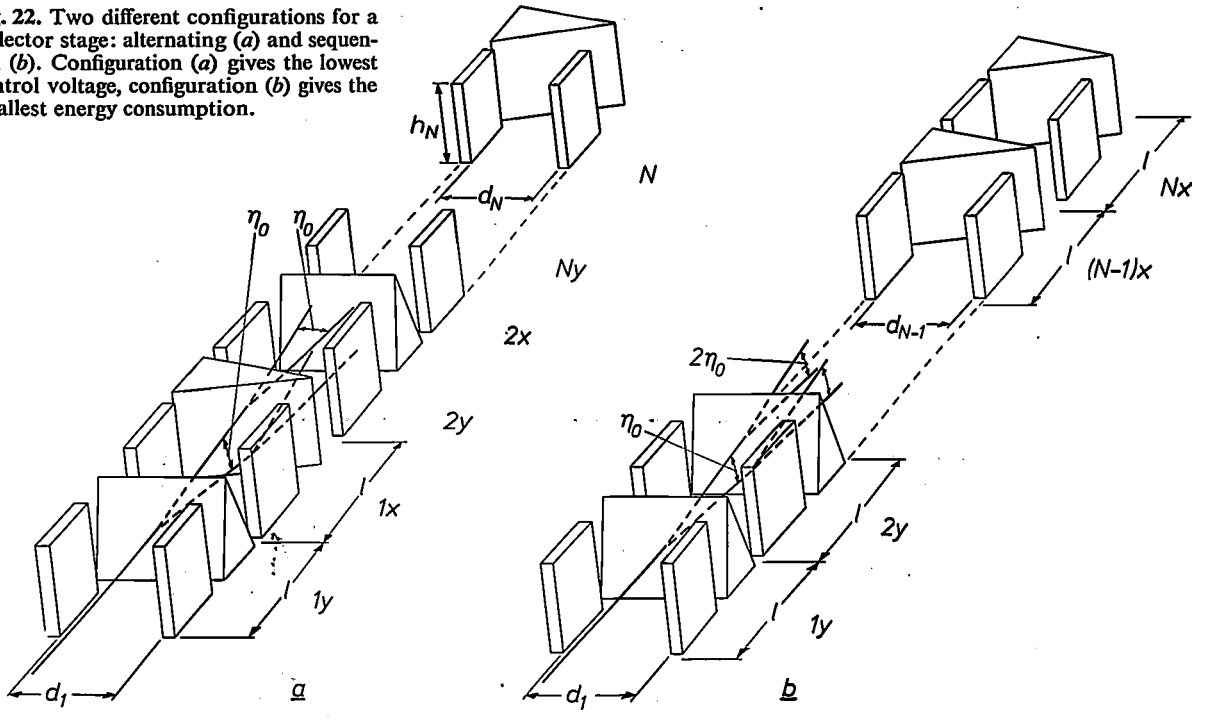


Fig. 22. Two different configurations for a deflector stage: alternating (*a*) and sequential (*b*). Configuration (*a*) gives the lowest control voltage, configuration (*b*) gives the smallest energy consumption.

a beam of this diameter ($d = d_1$), then the deflector devices for each successive stage must be larger. From fig. 22*a* we find for the width d_N and the height h_N of stage N:

$$d_N = d + 2^N (2l)\,\beta\lambda/d, \tag{10a}$$

$$h_N = d + 2^N (3l)\,\beta\lambda/d. \tag{10b}$$

The quantity β is related to the elementary splitting angle η_0 by $\beta = \eta_0 d/\lambda$; l is the length of a stage. For the second equation it is also assumed that $2^N \gg N$. If the electrodes are oriented as in the diagram and if we assume that the length l_p of the prism in each stage is small with respect to the electrode length l_e, we find that the voltages for the deflectors are independent of the length of a stage, and hence of the length of the deflector, while the power required increases with the electrode length.

The deflector can now be optimized with the aid of these equations, for example with respect to the parameters already mentioned: the control power or the control voltage. In the case of control power, a distinction can once again be made between the optimization of the complete deflector or of a particular stage. We shall consider the optimization of a stage, the simpler case, since optimization of the complete deflector requires a knowledge of the method of use, e.g. as to whether the stages all operate at the same frequency.

Making use of equations (10) and Kerr's law (1), and also of the expression for the energy E stored in a capacitor:

$$E = \tfrac{1}{2} CV^2, \tag{11}$$

we find that the energy P_N used in stage N is given by the relation

$$P_N = \tfrac{1}{4}\,\varepsilon\varepsilon_0 d_N h_N f/B, \tag{12}$$

where $\varepsilon\varepsilon_0$ is the dielectric constant and f the switching rate of the laser beam. Differentiating this equation with respect to d gives expressions relating to a minimum of energy consumption by the last stage of the deflector:

$$d_N = 2.84\,(R\beta\lambda l)^{0.5}, \tag{13a}$$

$$h_N = 3.48\,(R\beta\lambda l)^{0.5}, \tag{13b}$$

$$P_{N,\min} = 4.95\,\varepsilon\varepsilon_0\, R\beta\lambda l f/2\beta. \tag{13c}$$

The control voltage of this stage is given by

$$V_N = 2.84\,(R\beta\lambda/2B)^{0.5}. \tag{14}$$

Similarly, the geometry at optimum voltage can be characterized by the equations:

$$d_N = 2.82\,(2^N\,\beta\lambda l)^{0.5},$$

$$h_N = 3.53\,(2^N\,\beta\lambda l)^{0.5},$$

$$V_{N,\min} = 2.82\,(R\beta\lambda/2B)^{0.5}, \tag{15}$$

$$P_N = 5\,(\varepsilon\varepsilon_0/2B)\,R\beta\lambda f. \tag{16}$$

An identical procedure can be performed for the stage with the configuration of fig. 22*b*.

If we compare the separate expressions for the voltage and the energy consumption, we find that the configuration of fig. 22*b* is most suitable for minimum power and that the configuration of fig. 22*a* requires the lowest control voltage.

Summary. A useful method for deflecting a laser beam through a fixed angle has been developed at Philips Forschungslaboratorium Hamburg. The method uses a combination of a birefringent calcite prism and a Kerr cell filled with nitrobenzene. The beam follows the direction of the ordinary or the extraordinary ray, depending on the presence of voltage at the Kerr cell; this voltage has a value such that it changes the polarization plane of the beam by 90°. An experimental imaging system has been made with deflectors consisting of 20 deflector stages of this type — the largest that has been made contains 22 stages. This system gives an image of several m², consisting of 1024×1024 (i.e. 2^{20}) picture elements, with no granulation. The system is controlled by a computer. The Kerr cells used permit switching rates up to 500 kHz. The beam need not scan the screen line by line, but can jump directly to any other picture element. No laser light is then lost and the image can be viewed in a normally lighted room. One of the prototypes of such a system is already in use elsewhere in Germany. Other applications for the deflectors exist, e.g. in magneto-optic memories.

Transport of tungsten by the water cycle

From the earliest times in the development of lamps with tungsten filaments, a transport cycle was recognized whereby the tungsten was oxidized at a high temperature, under the influence of water vapour, to give a volatile compound that could be reduced to tungsten again at colder locations in the lamp.

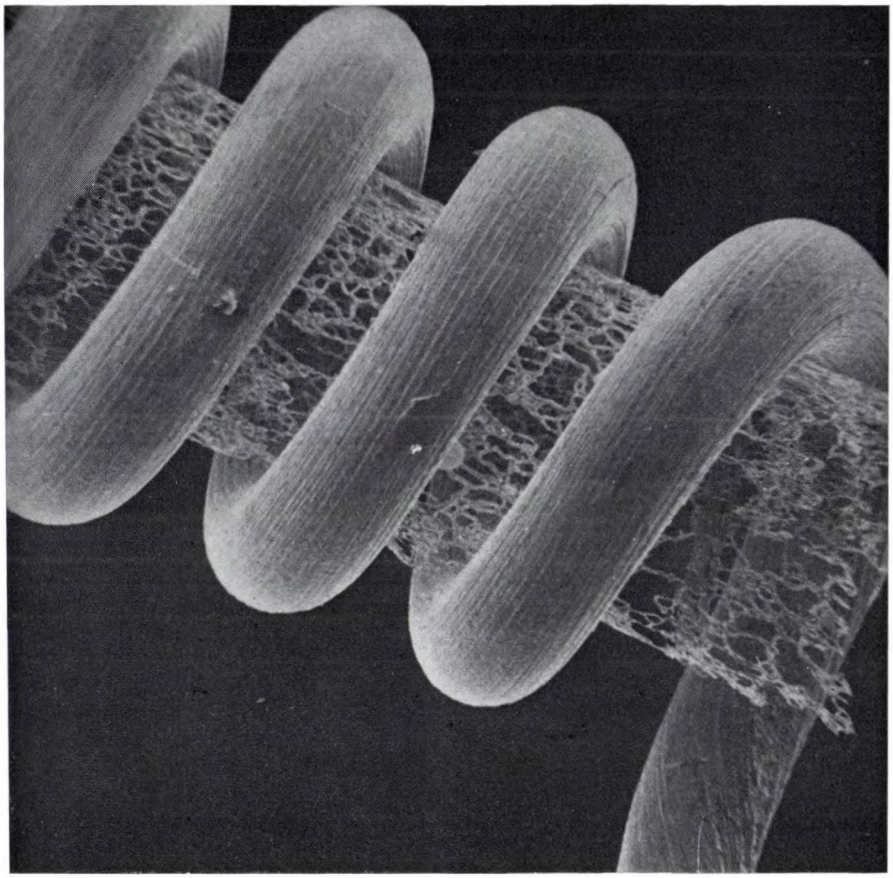


Fig. 1. Tungsten coil enclosing a lacy cylinder of tungsten-molybdenum alloy. This has developed during the setting of the coil in a moist atmosphere, which has caused the tungsten to be deposited on the grain boundaries of the molybdenum mandrel. The W-Mo alloy formed locally does not dissolve in the etchant used for removing the mandrel. (Magnification 300×.)

Although Irving Langmuir had postulated an explanation for this transport as far back as 1913 [1], it was to be another 50 years before sufficiently reliable thermodynamic data were available for the development of a more complete and convincing picture of the transport mechanism [2].

The tungsten is transported in the form of a volatile hydrated oxide. The transport reaction can be described by the equation

$$W + 4H_2O \rightleftarrows WO_3H_2O\uparrow + 3H_2.$$

This reaction moves in the appropriate direction in accordance with the law of mass action; the equilibrium depends on the ratio of the partial pressures of the water and the hydrogen. The equilibrium is also very sensitive to temperature, tending to the right with rising temperature: above 2000 °C the oxidation rates can be very rapid, and this applies for typical filament operation. At cooler parts of the system such as the envelope deposits of oxides or metal will occur. The composition of these deposits is again determined by the gas composition and the local temperature.

In the case of lamps this transport is a nuisance, since minute residues of water vapour can lead to erosion of the filament and darkening of the bulb. There are also other cases in which the tungsten

[1] I. Langmuir, Proc. Amer. Inst. Electr. Engrs. **32**, 1895, 1913.
[2] G. R. Belton and R. L. McCarron, J. phys. Chem. **68**, 1852, 1964.
F. H. R. Almer and P. Wiedijk, Z. anorg. allgem. Chemie **385**, 312, 1971.

transport can be troublesome. One example is the occurrence of 'insoluble mandrels' in the manufacture of tungsten filament coils. In the 'setting' of filaments wound on a molybdenum mandrel the presence of water in the gas mixture can make the conditions favourable for tungsten transport. Although in this case there is no temperature difference, any tungsten deposited on the molybdenum mandrel dissolves and forms an alloy at the surface, preferentially at the grain boundaries, since it is there that the chemical activity and the diffusion rate are highest. After etching, a thin lacy network remains that gives a striking and attractive 'memory' of the grain boundaries at the surface of the mandrel (*fig. 1*). This network has been found to contain 50% tungsten. In those cases in which the heat treatment is continued for longer, the individual meshes of the network gradually fill up, so that a full shell tubing is eventually produced. This is a nuisance in a production process, but the structures developed offer an elegant display of thermodynamic principles.

Fortunately, this transport reaction is not invariably undesirable. It is relatively easy to arrange the conditions so that a cohesive layer is deposited at acceptable rates [3]. This makes the water-transport cycle an alternative to other methods, such as the reduction of tungsten halides or carbonyls [4].

By using the water transport cycle it is possible to apply a thick and strongly adherent tungsten layer to substrates that do not oxidize at high temperatures in a water-vapour atmosphere. Molybdenum and nickel, like most of the higher-melting-point metals, can readily be coated, but not high-chromium alloys; these are too readily oxidized to be suitable.

The apparatus can be relatively simple, as for most chemical transport processes. The tungsten tube acting as the source and the substrate to be coated are placed in sequence inside an envelope. A stream of humidified hydrogen is passed through the tube that is held at the desired temperature by r.f. heating. To obtain a homogeneous and thick layer it is also desirable to heat the substrate. It is of essential importance here to choose the temperature and the gas flow correctly. If the conditions are not optimum, single-crystal tungsten dendrites may be produced (*fig. 2*); while these are aesthetically attractive, they are not very useful.

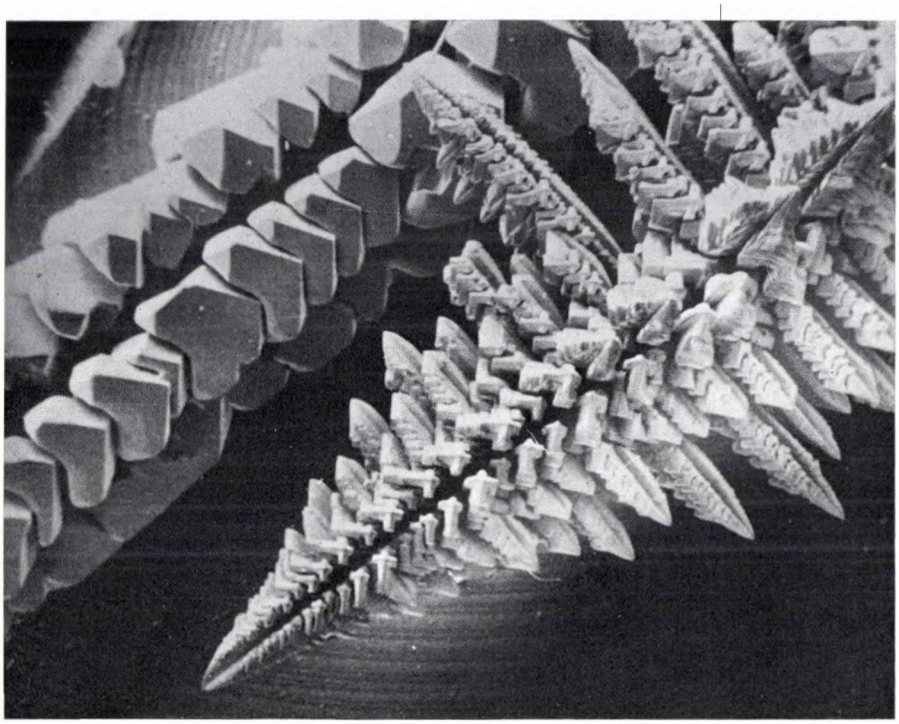


Fig. 2. Tungsten dendrites that have formed during the plating of tungsten under unsuitable conditions. These undesirable by-products can assume very attractive shapes. (Magnification 45×.)

[3] G. L. Davis and M. G. Charlton, High Temp. - high Press. **6**, 277, 1974.

[4] J. F. Berkeley, A. Brenner and W. E. Reid, Jr., J. Electrochem. Soc. **114**, 561, 1967.
F. W. Hoertel and J. G. Donaldson, U.S. Bur. Mines, Rep. Invest. No. 6915, 1967.
See also the articles 'Research on incandescent lamps', Philips tech. Rev. **35**, 295-306, 1975 (No. 11/12), with their quoted literature.

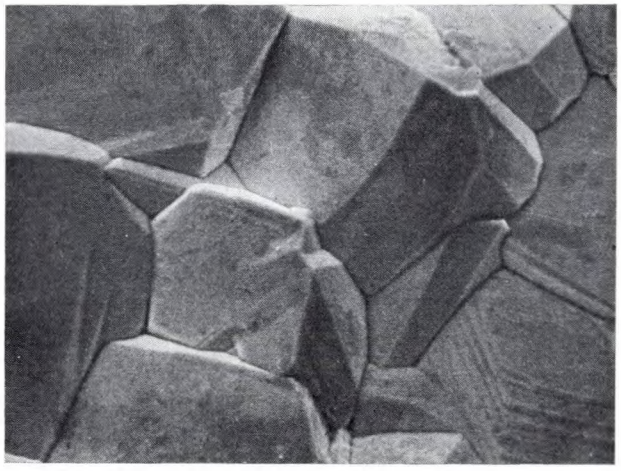


Fig. 3. The surface of a tungsten deposit obtained under optimum operating conditions for the water cycle. (Magnification 2500×.)

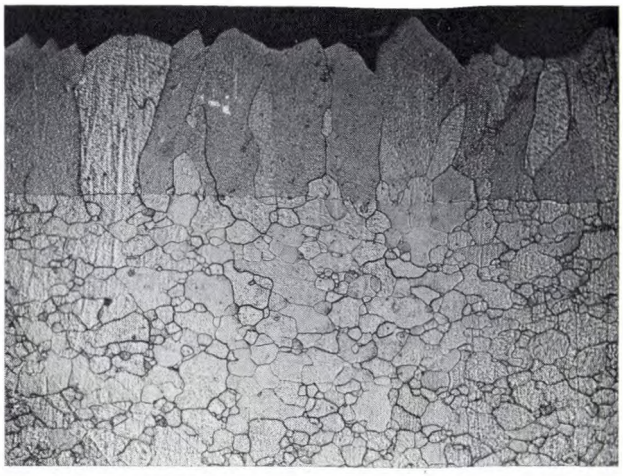


Fig. 4. Cross-section of a tungsten deposit produced on a molybdenum substrate (below) under optimum conditions. It can be seen clearly that the molybdenum grain boundaries continue into the tungsten layer. (Magnification 200×.)

With favourable conditions, tungsten layers of extremely high purity are obtained, apart from a small oxygen content (up to about 0.05%); these layers are free of porosity. As mentioned earlier, they adhere well to non-oxidizing substrates. The deposits do not recrystallize, even when they are heated to high temperatures. The surface of the coating is faceted (*fig. 3*), so that the surface gradually becomes rougher as the thickness of the coating is increased.

The crystals are deposited in a columnar structure, with the grain boundaries tending to run more or less normal to the surface of the substrate. An epitaxial layer can be obtained on a molybdenum substrate, and also on other suitable materials, as can clearly be seen in the cross-section shown in *fig. 4*.

G. L. Davis

Dr G. L. Davis is with Mullard Ltd, Mitcham, Surrey, England.

Intermetallic compounds; background and results of twenty years of research

J. H. N. van Vucht

The rich variety presented by the elements of the periodic system suggests a countless number of ways of finding new materials with new properties. This article reports on just such an exploration in the domain of the metals.

The motivations that prompted this investigation varied from a direct requirement from the manufacturer to a vague and intuitive feeling on the part of the investigator that a particular subject was going to become important to science and technology. Sometimes, an investigation that seemed to be petering out came suddenly to life again years later with the discovery of new sustenance.

Such a course of events is probably characteristic of all research: although it may appear to be rather random and unplanned, it does in fact have direction and purpose. In this particular sector the end results have included an increase in the $(BH)_{max}$ of magnetic materials to a value of 164 kJ/m³ (20.2×10^6 gauss oersted) — for a time the highest value attained anywhere — a breakthrough in the development of shape memories, and the discovery of a material, $LaNi_5$, whose astonishing capacity for storing hydrogen opens the way to a number of interesting potential applications.

Introduction

The search for substances with new properties

Anyone seeking to make new materials — in our case metallic materials — that have interesting new or greatly improved properties, will do well to reflect beforehand on the origin of these properties, both physical and chemical. He will then find it necessary to distinguish between *intrinsic* and *extrinsic* properties: the intrinsic properties are those attributable only to the nature of the atoms and their interrelationships (interaction, lattice structure), the extrinsic properties are those imposed on the material, deliberately or not, by deviations from the ideal structure.

Changes in both the intrinsic and extrinsic properties of a material can be brought about by physical and chemical methods. We shall assume that physical treatments such as deformations and quenching and their effects on material properties are sufficiently well known; we should like to take a closer look at the chemical methods here.

The chemical methods consist in bringing other elements into play. The resultant changes cannot always readily be distinguished as intrinsic or extrinsic, for when a second element is introduced there is the possibility of more than one phase being stable, which necessarily leads to phase boundaries. The extent, number and nature of such phase boundaries are open to variation, which leads to changes in extrinsic properties. If we ignore these possible secondary influences of the chemical method, in short if we look only at the purely intrinsic changes, we can then distinguish between three possibilities. This will be explained in more detail for systems consisting of two components [1].

In the first possibility, which is illustrated by the phase diagram in *fig. 1a*, it is energetically more advantageous for atoms A and B to keep away from each other and for each atom to surround itself with kindred atoms. At the temperatures where only the solid aggregation state is stable, and outside the regions in which A and B are mutually soluble — which may be extremely small — we always find a *mixture of two phases* that are in equilibrium (coexistence) with each other. The intrinsic properties can be thought of as arising from the superimposition of the intrinsic properties of each of the components.

A second possibility is illustrated by the phase diagram in fig. 1*b*, which relates to metals A and B whose chemical affinity for each other differs very little from their affinity for their own kind. Both elements have the same crystal lattice and each kind of atom occupies as many lattice points, in a random distribution, as

Dr J. H. N. van Vucht is with Philips Research Laboratories, Eindhoven.

corresponds to its composition percentage (*mixed crystal*). In this case the intrinsic properties change gradually from those of pure A to those of pure B. This need not happen monotonically and may be associated with a maximum or a minimum.

temperatures from a disordered to an ordered structure can have electrical, magnetic or mechanical consequences.

If we take two metals A and B that have a strong mutual preference as our starting elements, they will in

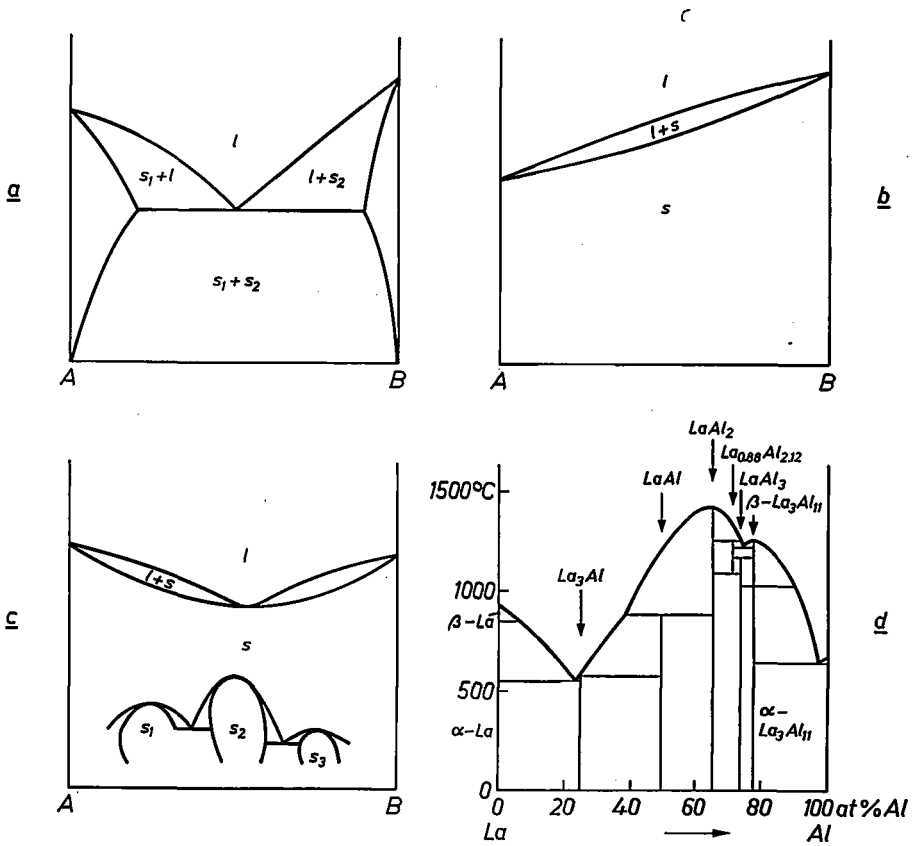


Fig. 1. *a*) Schematic phase diagram of a binary system whose atoms show a marked preference for their own kind. An example is the system Ag-Cu. *l* liquid, s_1 an A-rich solid phase and s_2 a B-rich solid phase. In the region $s_1 + s_2$ a mixture is stable that consists of two kinds of crystals, e.g. nearly pure Ag and nearly pure Cu crystals. *b*) Schematic phase diagram of a binary system whose atoms show no preference for being surrounded by atoms of their own kind and may equally be surrounded by atoms of a different kind, e.g. Ag and Au. In the region *s* all the crystals have the same lattice in which, depending on the composition, more or fewer A and B atoms are distributed in a disordered way among the lattice sites. *c*) Schematic phase diagram of a binary system whose atoms show a slight preference for being surrounded by atoms of the other kind, e.g. Cu and Au. At low temperatures new lattices are formed (s_1, s_2 and s_3) which roughly resemble those of the pure components but possess a different periodicity. At high temperatures the weak energetic preference for being surrounded by atoms of a different kind is overcome by the effect of entropy. *d*) Phase diagram of the system La-Al [2], as an example of a system that is rich in intermetallic compounds, with new physical and chemical properties, because of a marked preference for being surrounded by atoms of the other kind. This diagram demonstrates the very different behaviour of the various compounds on a change in temperature: the highest melting point is that of $LaAl_2$; at a temperature increase LaAl decomposes into another solid and the liquid, and at a fall in the temperature $La_{0.88}Al_{2.12}$ decomposes into two other solids.

The third possibility is explained in fig. 1*c* and *d*.

If the constituent atoms A and B show a slight preference for the neighbourhood of the other kind, a completely disordered mixture will occur at high temperatures, but at low temperatures a certain ordering will take place. A *new* phase then arises with a *new* lattice (although closely related to that of A and B) and with new properties. A transition of this kind at low

general show a strong inclination to form entirely new types of lattice with varied configurations of A and B atoms, depending on their mixing ratio. Thus, in the La-Al system, whose phase diagram is given in fig. 1*d*,

[1] A detailed treatment of phase diagrams is given in J. L. Meijering, Philips tech. Rev. **26**, 12 and 52, 1965.

[2] K. H. J. Buschow and J. H. N. van Vucht, Philips Res. Repts. **22**, 233, 1967.

six intermetallic compounds are formed with lattices that differ from each other and from those of the starting elements. All have their own intrinsic properties.

It will be clear from the foregoing that the investigator looking for new properties will prefer to look for such new lattices, because there is a greater probability of finding sharp changes in properties. A new lattice of this kind may form the starting point for further optimization that can lead to the desired property. This optimization can be carried out chemically, e.g. by making suitable mixed crystals, or physically, e.g. by introducing the appropriate lattice defects.

In this article we shall take examples from widely different fields — getters, superconductors, permanent magnets, shape memories and hydrogen storage devices — to illustrate how in a number of cases the adoption of the strategy described here has led to interesting and even spectacular scientific and technical results.

Other research tools

Work has been done at Philips Research Laboratories for a number of years on a wide variety of metallic systems, with a variety of suitably adapted methods. The preparatory methods are all designed to minimize the incorporation of foreign elements in starting materials of the highest possible purity. Thermal analysis is used to obtain a knowledge of the exact details of the phase diagrams. X-ray analysis is also used, where necessary at temperatures ranging from 5 K to 2000 K. At a somewhat more advanced stage X-ray analysis is employed to determine the precise atomic structure of the compounds [3].

Measuring the properties of the compounds obtained involves the use of a great diversity of techniques. In the early stages of searching in a wide area it is desirable to obtain a quick impression; intensive and accurate measurements will only be carried out if real promise appears.

We have been able to carry out these measurements in a wide temperature range and also, in recent years, in a large range of pressures [4]. We can for example measure magnetic properties and electrical conductivity at temperatures up to 2000 K and at pressures up to 10^5 bars, though it must be admitted that not all combinations of these measurement conditions are yet possible.

About twenty years ago research on intermetallic compounds was mainly directed towards materials that could be used as getters in small thermionic valves. We shall deal with this work first because of the many connections between it and the subjects of later investigations.

Getters

In thermionic valves a getter is generally introduced for the purpose of chemically binding the gases released after the evacuated valve has been sealed off. The getter metal most commonly used is barium, which is vaporized by heating it in the sealed-off valve, so that it condenses in the form of a thin metallic film on the valve wall. In this state the barium is highly active and absorbs the residual gases, notably hydrogen, oxygen, nitrogen, water vapour and carbon dioxide and to a lesser extent hydrocarbons and carbon monoxide.

This type of getter could not however be used in the miniature valves made for hearing aids, because the metal film causes short circuits and unwanted capacitances.

In 1945 a non-evaporating getter became available that was found to be ideally suited for binding the principal residual gases, hydrogen and oxygen, in these miniature valves. This getter was made by sintering thorium powder together with a powder of an aluminium alloy and 'Cer-Misch-Metall' (an alloy of rare-earth metals with cerium as the main constituent). The getter, in the form of a fine powder, is sintered to a component of the valve and is readily activated by heating it, e.g. electrically, after the valve has been sealed off.

In an endeavour to explain the operation of this getter we made an extensive study of the ternary system thorium-cerium-aluminium. We found that the getter is an intermetallic compound with the crystal lattice of Th_2Al (*fig. 2*), in which the cerium atoms are randomly distributed over the thorium sites. It was found that there are interstices between the thorium atoms into which hydrogen atoms easily fit. The pure metal thorium also contains such interstices, and this metal can likewise be used for binding hydrogen, but Th_2Al is more suitable for two reasons. Firstly, hydrogen diffuses much faster in the Th_2Al lattice than in the Th

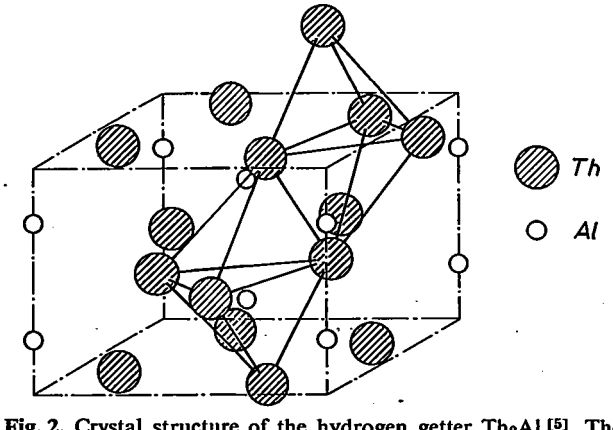


Fig. 2. Crystal structure of the hydrogen getter Th_2Al [5]. The hydrogen atoms prefer to occupy the tetrahedra formed by the Th atoms, two of which are shown in the figure.

lattice, partly because the hydride lattice of Th_2Al that is formed contains more vacancies than that of thorium. Secondly, it is easier to make a large reactive surface from Th_2Al because, unlike thorium, it is very brittle.

What exactly is the function of the cerium? Unlike hydrogen, oxygen will only diffuse reasonably quickly into a metal lattice like that of Th_2Al at a very high temperature. The first reaction with oxygen to occur at the outer surface leads to the formation of an oxide layer, which has a sealing effect for pure Th_2Al, and rules out the possibility of any further reaction with oxygen, or even with hydrogen. Cerium ensures that the oxide coating on the getter remains permeable, so that the rate of reaction with oxygen and hydrogen remains high.

This is presumably a consequence of the fact that the element cerium can be either trivalent or tetravalent. When cerium is oxidized, Ce_2O_3 forms first and then CeO_2. The Ce_2O_3 lattice may be regarded as a somewhat more rarefied CeO_2 lattice containing an ordered arrangement of oxygen vacancies. The presence of these vacancies allows oxygen to diffuse rapidly through the Ce_2O_3 layer formed on oxidation on the metal side. The CeO_2 layer formed at the oxygen side of the Ce_2O_3 is much less bulky than the Ce_2O_3 and therefore does not form a complete seal.

We therefore see that there is a sudden marked improvement in intrinsic getter properties when the thorium lattice is converted into the lattice of Th_2Al, and that a further modification of that lattice by dissolving cerium atoms in it gives a further improvement — an optimum result is achieved by substituting cerium atoms for about one in every eight thorium atoms.

Superconductors

The property known as superconductivity, in which the electrical resistance drops to zero at very low temperatures, is found in many metals, intermetallic compounds and alloys. For the elements, the critical temperature T_c (transition temperature), the maximum temperature at which superconductivity occurs, is highest in niobium: 9 K. It would be useful to have superconductors with a transition temperature of at least 20 K, since it would then be possible in principle to use liquid hydrogen for cooling instead of liquid helium.

It has been found that the value of the transition temperature depends on the average concentration of the valence electrons. If tetravalent zirconium atoms are substituted for pentavalent niobium atoms in the niobium lattice, the result is to increase the concentration of the valence electrons, and the transition temperature can be raised from 9.2 K to 10.8 K. In a similar way the substitution of tetravalent tin atoms for trivalent indium atoms at arbitrary sites in the $LaIn_3$ lattice, bringing the concentration of valence electrons to more than 3.5 electrons per atom, results in a marked increase in the transition temperature (*fig. 3*).

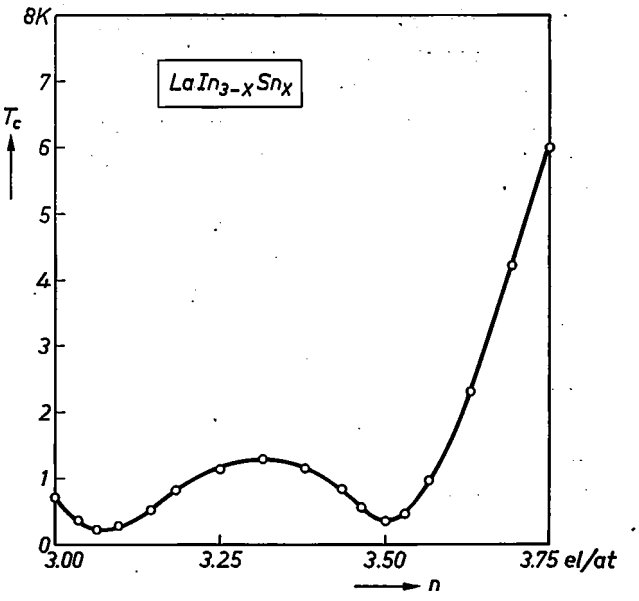


Fig. 3. The transition temperature T_c, i.e. the highest temperature at which superconductivity occurs, as a function of the average number of valence electrons n in the $LaIn_{3-x}Sn_x$ lattice [6].

A marked improvement to 18 K was achieved with the intermetallic compound Nb_3Sn, which has a characteristically different lattice (*fig. 4*). A characteristic feature of the configuration is that the rows of Nb atoms are perpendicular to each other, and at a 10% smaller spacing than in pure niobium. The reduction of the distance between the niobium atoms is evidently made possible by the stabilizing effect of the Sn atoms, which are situated in the interstices formed by the network of rows. The value of the transition temperature does not appear to depend primarily on the nature of the stabilizing atoms. A lattice of this type, with a composition $Nb_3(Al_{0.75}Ge_{0.25})$ for example, has a transition temperature of slightly more than 20 K [7].

A complication, however, is the fact that it is not always easy to prepare such a lattice with the correct

[3] See P. B. Braun and A. J. van Bommel, Philips tech. Rev. 22, 126, 1960/61.

[4] The method used will be the subject of a forthcoming article in this journal.

[5] J. H. N. van Vucht, Interaction of Th_2Al and related getters with hydrogen, thesis, Eindhoven 1963.

[6] E. E. Havinga and M. H. van Maaren, Physics Repts. **10C**, 107, 1974, in particular fig. 28.

[7] G. Arrhenius, E. Corenzwit, R. Fitzgerald, G. W. Hull, Jr., H. L. Luo, B. T. Matthias and W. H. Zachariasen, Proc. Nat. Acad. Sci. U.S.A. **61**, 621, 1968.

stoichiometric composition and with the desired degree of order, so that the result obtained is not the optimum. Recently, however, this has been achieved in the preparation of Nb_3Ge, and this has brought the critical temperature to the previously unequalled value of 23.2 K [8].

It should be noted in passing that the information we obtained in that study about the phase diagram of the Nb-Sn system enabled us to understand why it was never possible to make Nb_3Sn from the elements at temperatures below 950 °C, even after protracted heating [9]. The explanation lies in the formation of the compounds Nb_6Sn_5 and $NbSn_2$, which result in layers that form a seal. Above 930 °C these compounds are not stable. With the knowledge thus obtained we succeeded in devising a process in which Nb_3Sn can be formed fast enough to produce superconducting coils.

Permanent magnets

We have already described how our research on getters led to a better understanding of the rare-earth metals, and of cerium in particular. These rare-earth metals attract attention much more for their magnetic than for their gettering properties.

In the series of rare-earth metals arranged in order of atomic number, also known as the lanthanide series, the electron shells are not filled in the outer shell but in the deeper 4f shell; this has the result that all these elements are trivalent. Since in filling the 4f shell in the order mentioned electrons with a parallel spin are added first and only afterwards electrons of opposite spin, the spin S per atom first increases and then drops to zero. The coupling of the orbital momentum L takes place in the first half of the series by the subtraction $L - S$, in the second half by the addition $L + S$. The resulting magnetic moments are shown in *fig. 5*.

Quite apart from these effects there is a reduction in the atomic radius, called the lanthanide contraction, which is caused by the increase in the nuclear charge. This means that we have here a series of chemically almost identical elements that differ in atomic radius and in magnetic behaviour.

The 4f shell that 'carries the magnetism' is usually unaffected during the formation of compounds, and this makes it interesting to study the magnetic interaction of the rare-earth atoms with non-magnetic atoms in the same lattice and also with magnetic atoms of other elements, such as iron, cobalt, manganese or nickel. In the case of the system Sm-Co the combination of magnetic atoms of different kinds in one lattice led to a breakthrough in permanent-magnet technology.

Material for permanent magnets should have a high saturation magnetization and a high crystal anisotropy. The saturation magnetization is not particularly high in the system Sm-Co; cobalt has a relatively large magnetic moment per atom, but that of samarium is relatively small owing to the subtraction of spin and orbital momentum (fig. 5). However, the hexagonal lattice of $SmCo_5$ in particular does have a very high crystal anisotropy (*fig. 6*), which gives this material its special value.

This journal has already reported on the race to achieve ever higher values of $(BH)_{max}$, which culminated in a value of 164 kJ/m³ (20.2×10^6 gauss oersted) for $SmCo_5$ sintered in a special way [10].

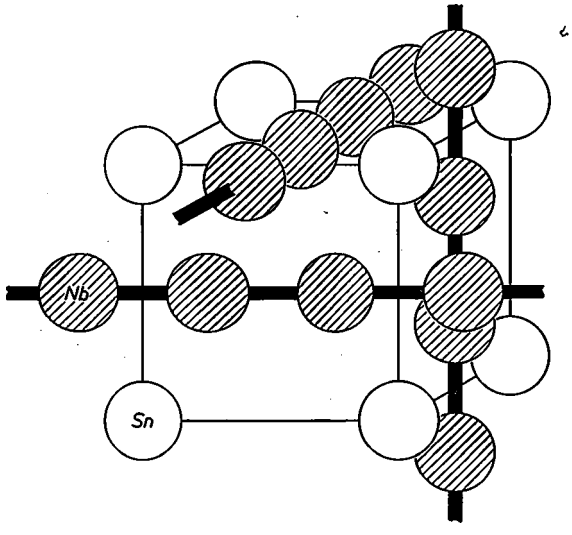


Fig. 4. Diagram of the crystal lattice of Nb_3Sn, in which three rows of Nb atoms at right angles to one another are connected by lines. In this material the distance between the Nb atoms is 10% shorter than in the pure niobium lattice, which has a favourable effect on the height of the transition temperature (18 K in Nb_3Sn as against 9.2 K in pure niobium).

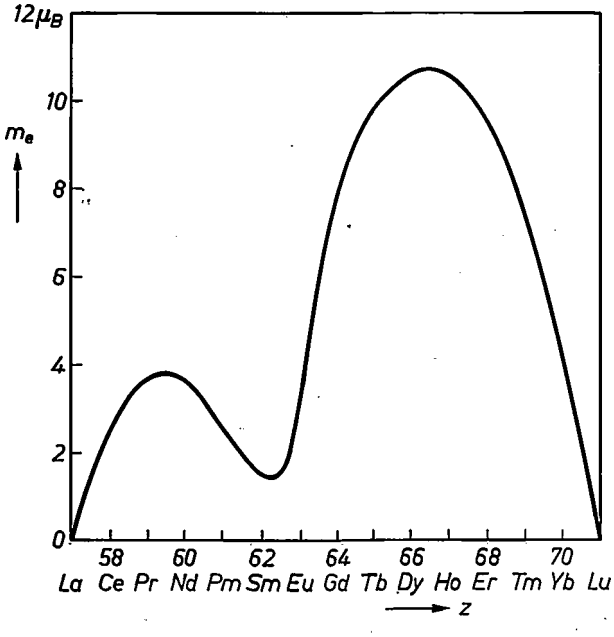


Fig. 5. The effective magnetic moment m_e per atom as a function of the atomic number z in the lanthanide series.

In 1972 a $(BH)_{max}$ of 207 kJ/m^3 was attained by replacing some of the samarium atoms in the $SmCo_5$ lattice by praseodymium atoms, which have a higher magnetic moment and hardly affect the crystal anisotropy on substitution. Not so long afterwards this record was broken by a group who succeeded in growing small single crystals of $SmCo_5$ in which they also replaced some of the samarium atoms by cobalt atoms, which have a higher magnetic moment. The magnetic moment of these substituted cobalt atoms aligns itself in the lattice parallel to that of the other cobalt atoms in the lattice. This effect, taken with the fact that the density and orientation of the magnetic atoms in a single crystal are optimum, led to a $(BH)_{max}$ of 254 kJ/m^3 [11].

The significance of this breakthrough lies in the new scope it allows to the designers of various kinds of equipment since the magnets can now be made so small and light.

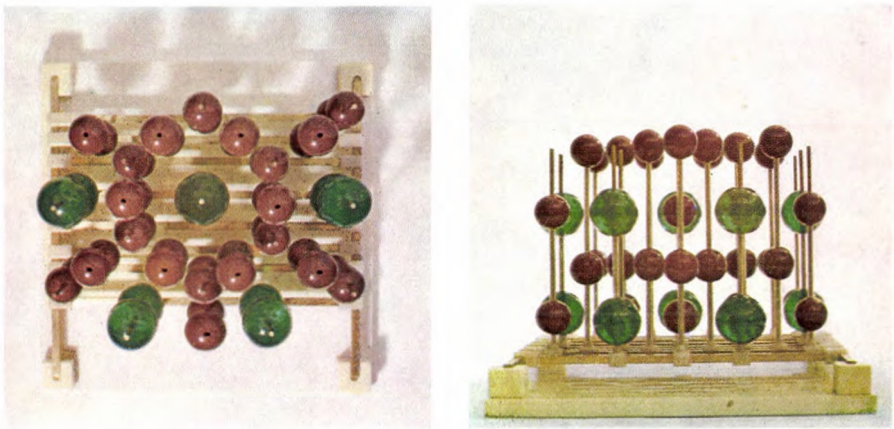


Fig. 6. Model of the hexagonal lattice $SmCo_5$ (Sm green, Co brown) illustrating how the pattern of the structure in two directions at right angles is entirely different. This explains the high crystal anisotropy of this material.

Polytypism of close-packed AB_3 compounds

The fact that the atomic radius in the lanthanide series changes while the number of valence electrons remains constant gives the investigator the chance to study the geometrical factors that play a role in the stability of compound lattices, since the concentration of the valence electrons, which also has an effect, remains the same in all cases.

As a direct offshoot of the work on Th-Ce-Al getters we investigated the R-Al systems (where R stands for an element from the lanthanide series). One of the interesting results of this work was the discovery that a tri-aluminide occurs in nearly all these systems, but not with the same crystal structure.

It was found that at the beginning and end of the lanthanide series the structure of the RAl_3 is very simple: from $LaAl_3$ to $GdAl_3$ there is a hexagonal close-packed structure, and from $ErAl_3$ the cubic close-packed structure is found (*fig. 7*), but between these there are unknown complex structures.

We assumed that we were concerned here with *polytypes*. The term polytypes refers to lattices that are identical in two dimensions and differ from each other only in one dimension [12].

The hexagonal and the cubic close-packed structures may be regarded as the two limiting cases of a series of polytypes. They can be thought of as built up from the same two-dimensional layers, which differ from each other in the third dimension because of a different method of stacking.

If we look at one atomic position in such a layer (position *A*), we see that in the cubic close-packed structure this atomic position in the adjacent layer has been displaced by 1/3 of the diagonal of the unit cell (position *B*), then again by 1/3 of the diagonal in the next layer, (position *C*), and that in the third layer it returns to position *A*. In the hexagonal close-packed structure the return to the first position takes place

[8] L. R. Testardi, J. H. Wernick and W. A. Royer, Solid State Comm. **15**, 1, 1974.

[9] J. H. N. van Vucht, D. J. van Ooijen and H. A. C. M. Bruning, Philips Res. Repts. **20**, 136, 1965.

[10] See for example:
K. H. J. Buschow, W. Luiten, P. A. Naastepad and F. F. Westendorp, Philips tech. Rev. **29**, 336, 1968;
F. F. Westendorp and K. H. J. Buschow, Solid State Comm. **7**, 639, 1969;
H. Zijlstra, Philips tech. Rev. **34**, 193, 1974.

[11] A. S. Ermolenko, A. V. Korolev and Ya. S. Shur, JETP Letters **17**, 359, 1973.

[12] This effect has also been observed in silicon carbide, for example; see A. H. Gomes de Mesquita, Philips tech. Rev. **30**, 36, 1969.

immediately after the first displacement by 1/3 of the diagonal. Thus, in the cubic close-packed structure the stacking order is *ABCABC* . . . , and in the hexagonal structure the order is *ABAB* . . . .

For our purposes a different notation is useful. In the hexagonal close-packed structure each layer lies symmetrically between its two neighbouring layers, whereas in the cubic case its situation is asymmetrical. In our notation, *h* (for 'hexagonal') denotes each layer that lies symmetrically between its neighbouring layers and *c* (for 'cubic') denotes each layer in an asymmetrical position. With this notation the transition from cubic close packing (*cc* . . .) to hexagonal close packing (*hh* . . .) that takes place in the lanthanide series can be symbolically expressed by *cc* . . . , *hcchchcchc* . . , *hchc* . . . , *hhchhc* . . . and *hh* (*fig. 8*).

Why does the material show a growing preference for a more hexagonal polytype as the atomic radius increases (and the atomic number decreases)? In all layers the large minority atoms R are surrounded by six smaller Al atoms (*fig. 9*). In the hexagonal packing the atoms located at identical positions are separated from each other by only one layer (compared with two layers in the cubic case). As the R atoms increase in size, the small Al atoms can more easily make room for them in the interstices. In the hexagonal packing the dispersal of the small atoms makes room for the large atoms in *both* neighbouring layers, leading to a closer packing. In the cubic case the advantage of such dispersal for the large atoms in the one neighbouring layer turns into a disadvantage for the large atoms in the other. The larger the R atoms the greater the gain in free energy of the closer packing thus achieved, and hence the increasing preference for the more hexagonal polytype.

Research on alkaline-earth-lead compounds and other materials [14] has shown that this is a general rule in crystal lattices with AB_3 compounds built up on the close-packed principle. In these compounds it is of course also necessary to take into account the effect of the electron concentration as well as that of the geometrical factors.

The knowledge thus obtained [15] enabled us to predict that the crystal structure of $Ba_5(Pb_{12},Tl_3)$ would show the stacking characteristic of a 5/7 hexagonal structure, i.e. the stacking *hhchhhc* (*fig. 10*). This prediction was confirmed by X-ray analysis of powders [16].

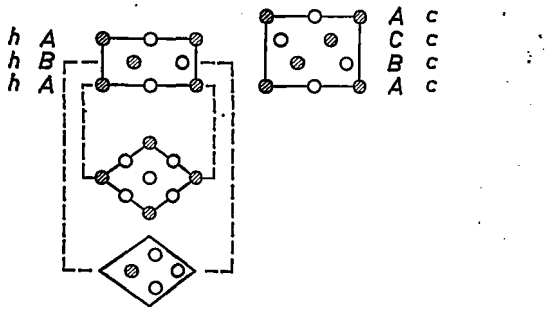


Fig. 7. Vertical cross-section (upper left) and two horizontal cross-sections (lower left) of the hexagonal unit cell of a close-packed AB_3 compound. The corresponding cubic stacking is illustrated on the right (only in vertical cross-section). In the stacking of the layers each atom is always displaced by 1/3 of the diagonal. The cubic stacking arises from successive displacements in the same direction, so that the original position is resumed in the third neighbouring layer. In the hexagonal case a zigzag progress is followed, and after one displacement through 1/3 of the diagonal the original position is resumed again. The notation often used for the cubic stacking is *ABCABC* . . . , and for the hexagonal stacking *ABAB* . . . In the hexagonal case each layer is symmetrical with respect to the neighbouring layers; in the cubic case it is asymmetrical. Calling the symmetrial position *h* and the asymmetric position *c*, the hexagonal stacking can be denoted by: *hh* . . . , and the cubic by *cc* . . . The value of this notation is clear from fig. 8.

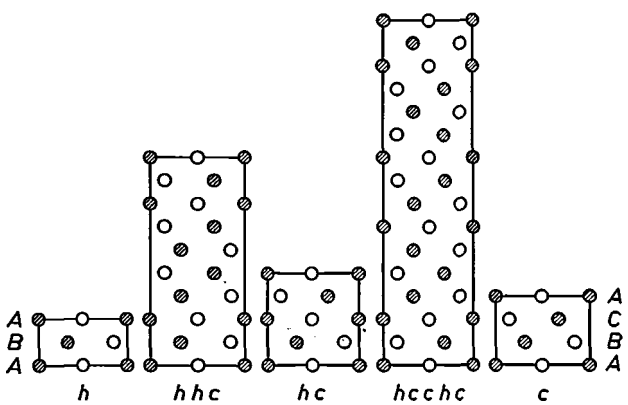


Fig. 8. Schematic representation of polytypes occurring in trialuminides of rare-earth metals [13].

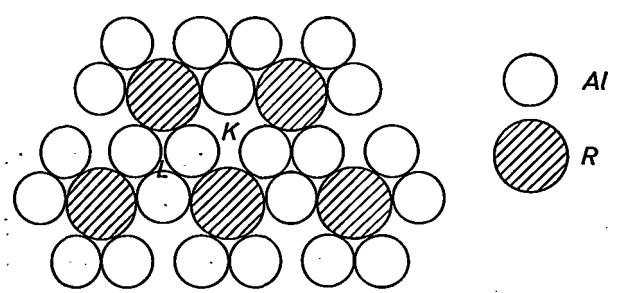


Fig. 9. A layer of a close-packed lattice of an RAl_3 compound, with R atoms that are large compared with the Al atoms. At the position *K* in the hexagonal case *both* R atoms of the identical neighbouring layers above and below can use the space left vacant by the displacement of the Al atoms. In the cubic case this only applies for the R atom of one of the layers; the R atom of the other layer is then located exactly at *L*, where the Al atoms have been crowded together by the dispersal around *K*. This prevents the formation of an optimum close-packed structure. The larger the R atoms are compared with the Al atoms the greater the tendency towards the formation of a hexagonal close-packed structure.

[13] A more detailed description is given in J. H. N. van Vucht and K. H. J. Buschow, J. less-common Met. **10**, 98, 1965.
[14] J. H. N. van Vucht, J. less-common Met. **11**, 308, 1966.
[15] E. E. Havinga, J. H. N. van Vucht and K. H. J. Buschow, Philips Res. Repts. **24**, 407, 1969.
[16] E. E. Havinga and J. H. N. van Vucht, Acta cryst. B **26**, 653, 1970.

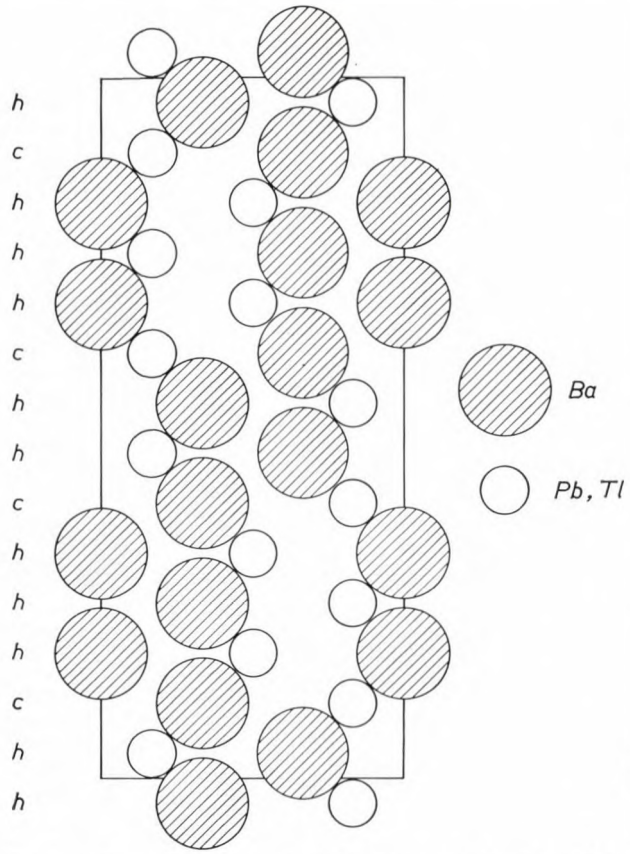


Fig. 10. Schematic representation of the structure of $Ba_5(Pb_{12},Tl_3)$.

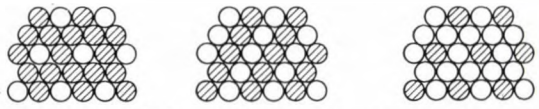


Fig. 11. Illustrating the relationship between close-packed AB lattices (centre) and AB_3 lattices (left and right), which applies for both the cubic and the hexagonal forms.

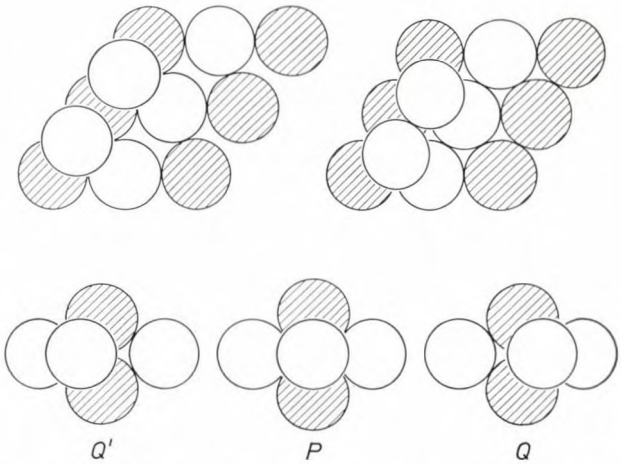


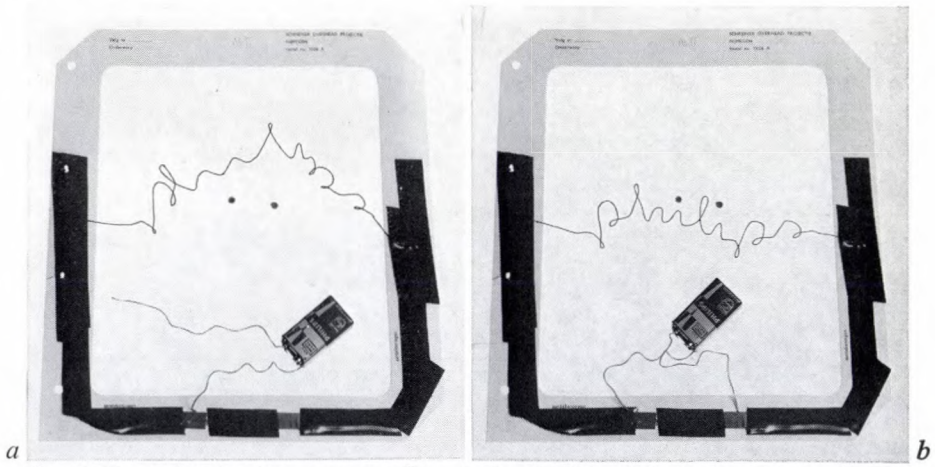
Fig. 12. Illustrating the transition from the ideal close-packed AB lattice (both cubic and hexagonal) to the less closely packed structure of the CsCl type (left), which provides the explanation for the shape-memory effect of a large number of alloys. *Q'* and *Q* are positions of the close-packed AB lattice (the upper atom lies in an interstice formed by the three lower ones), and *P* is a position of the lattice of the less closely packed type (the upper atom lies on a 'saddle point' of two atoms beneath it).

Shape memories

Close-packed AB compounds are related to close-packed AB_3 compounds, as illustrated in *fig. 11*. In compounds of the AB type it is also possible to imagine a series with a pure hexagonal close-packed structure at one end and a pure cubic close-packed structure at the other, with all kinds of polytypes in between. The search for such polytypes brought the intermetallic compound NiTi to our notice, a substance with a remarkable property, which was not understood at the time: shape memory.

AB compounds show more often than AB_3 compounds the tendency to change at high temperatures to a less close packing (*fig. 12*), in particular to structures of the CsCl type (or of the tungsten type in the case of elements). In NiTi the temperature at which the transition takes place from the closest packing to the less closely packed CsCl structure is only slightly above room temperature. It seemed an obvious assumption that this — diffusionless — transition was related to the memory effect observed.

Fig. 13 demonstrates the operation of a shape memory. Fig. 13*a* shows a wire of NiTi, which was originally shaped in the name of our firm and then deformed beyond recognition at room temperature.



a *b*

Fig. 13. Illustrating the shape-memory effect. A wire of NiTi, originally shaped to produce the name Philips, was deformed beyond recognition at room temperature (*a*); after the temperature was raised by passing an electric current through the wire, the wire returned to its former shape (*b*).

In fig. 13*b* it can be seen that when the temperature is raised the name Philips re-emerges; one indication of how accurately the original shape is recovered is that the dots in fig. 13*a* appear exactly above the i's.

In its closest-packed structure NiTi at room temperature is very ductile, and above the transition temperature of 60 °C it is hard, brittle and difficult to deform. Deformations applied at room temperature disappear, as we have shown, spontaneously and immediately when the temperature is raised above the transition temperature, and after cooling the metal returns exactly to the shape it had before the deformation: it has 'remembered' its former shape.

This spectacular property has found small-scale application in NiTi clamping rings, and many patents have been taken out for other applications varying from self-deploying satellite antennas to an application as a deformable wheel, capable of supplying energy from a fairly small temperature difference between two heat reservoirs.

Explanation of the shape-memory effect

Let us now try to picture the atomic events taking place during the structural transition. This can best be done by starting from the situation above the transition temperature (the CsCl lattice). The atoms of each layer lie on the 'saddle points' of the next neighbouring layer (fig. 12). When the temperature is lowered the structure changes to the most closely packed stacking (hexagonal, cubic or an intermediate form). When this happens the rows inside a layer move in such a way that the diabolo-shaped interstices change into triangular interstices. There is also a change in the relative position of the layers as a whole: the atoms of the one neighbouring layer then no longer lie on the saddle points of the other, as they did in fig. 12 (position P), but at the interstices (position Q).

At the transition temperature the two positions are thermodynamically equally stable: no activation free energy is necessary for changing from the one position to the other. At a somewhat lower temperature (when position Q is thermodynamically only very slightly more advantageous than position P) a force exerted on the lattice to deform it can very easily cause a relative displacement of the illustrated atomic planes. The atoms then jump from one interstitial position Q' (equivalent to Q) via a P position to the interstitial position Q.

In the case of chemically *symmetrical* triangular interstices the process may repeat itself from the new interstice via the two possible saddle positions that now become available in the direction of the strain. This explains the familiar fact that at the temperature of the transition from the W lattice to the cubic or hexagonal close-packed structure pure metals like iron or zirconium are very easily deformable.

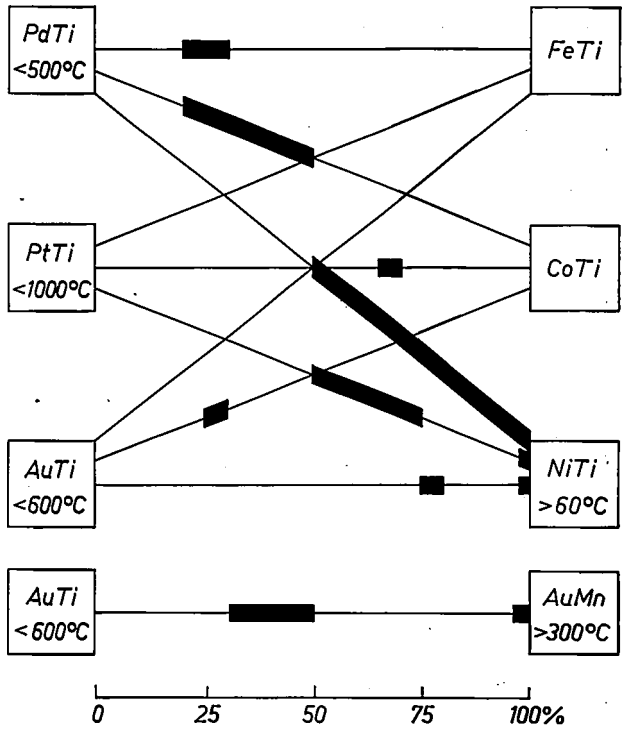


Fig. 14. Nomogram of a number of mixed systems of AB compounds that show the close-packed structure at room temperature (left) and AB compounds that have the CsCl structure at the same temperature. Compounds of the first type are ductile under these conditions, those of the second type are difficult to deform. Mixing on an atomic scale causes an abnormally high ductility in the concentration region (thick black lines), and this ductility is associated with the occurrence of a shape-memory effect. The explanation lies in the fact that the temperature at which a transition tends to take place from one type of stacking to the other (in some cases this lies below absolute zero) is shifted to the region just above room temperature on mixing.

These pure metals, however, do not give a memory effect. For this to be possible it is at least necessary that the atoms of a layer are able to 'feel' where they have come from after a displacement. If the layers are built up from *asymmetrical* interstices, this requirement is met. In AB compounds like NiTi (fig. 12) the atoms of one kind are found in rows that are in turn surrounded by rows of the other kind inside and outside the plane. If, before the deformation (at room temperature), atoms of the one kind are situated at position Q' and after the deformation at position Q, then during the deformation they are displaced across the saddle points in the rows of the other atoms. On a further displacement the atoms have to be moved across saddle points of a *different kind*, which have a higher activation free energy since they are also formed by an atom of their own kind. We therefore assume that for each shearing plane the deformation remains confined to this one 'saddle jump'. (Macroscopically the NiTi object may still have undergone a considerable change of shape.)

The original high-temperature position on the saddle point can clearly be distinguished from the two other possibilities; when the temperature is raised to above the transition point the atoms again return to this saddle point.

This picture of the atomic events taking place during such a transition made it clear that this property could not be unique to NiTi, and in 1970 it led us to the discovery of 23 previously unknown alloys with this shape-memory effect (*fig. 14*). We were even able to demonstrate that the copper-tin system, which has been studied since ancient times, also has a shape-memory effect in the composition Cu_3Sn. We do not know, however, whether this effect has ever been observed previously.

Storage of hydrogen

We have already dealt with the gettering properties of some of the intermetallic compounds of rare-earth metals, and have discussed the suitability of $SmCo_5$ as a material for permanent magnets. Reseach into these intermetallic compounds has yielded yet another surprising result.

Under the influence of the atmosphere, the coercive force of $SmCo_5$ and hence its $(BH)_{max}$ decrease noticeably. Investigation of this effect showed that it was associated with the hydrogen atoms, and the following new and highly important properties of $SmCo_5$ were discovered [17]. Firstly, the material, unlike an ordinary getter, very readily reacts with molecular hydrogen, even at room temperature; secondly, it releases the hydrogen again very easily; and thirdly, both the take-up and release of hydrogen take place in considerable quantities. In fact, $SmCo_5$ is a 'getter' with a high equilibrium pressure, somewhat higher than 1 bar.

For hydrogen taken up under high pressure to be released again it is therefore sufficient to lower the pressure to atmospheric pressure. With an ordinary getter the object is the permanent 'removal' of the hydrogen; any release of hydrogen is only permissible under rather drastic conditions. This newly discovered property, however, allows the hydrogen to be easily stored temporarily.

The next step was to determine how other compounds whose crystal structure is isomorphous with that of $SmCo_5$ would behave as a hydrogen-storage medium. It was found that the most suitable compound in this respect is $LaNi_5$ [18].

[17] H. Zijlstra and F. F. Westendorp, Solid State Comm. 7, 857, 1969.

[18] J. H. N. van Vucht, F. A. Kuijpers and H. A. C. M. Bruning, Philips Res. Repts. 25, 133, 1970.

The thermodynamic properties of the system $LaNi_5$-H_2 are indicated in *fig. 15*. At room temperature and a pressure of about 2.5 bars hydrogen can be absorbed until the composition $LaNi_5H_{6.7}$ is reached. This reaction is reversible, which means that if the hydrogen pressure is allowed to drop to less than two atmospheres — there is a slight hysteresis effect — the hydride form decomposes to the original metal, with the release of hydrogen.

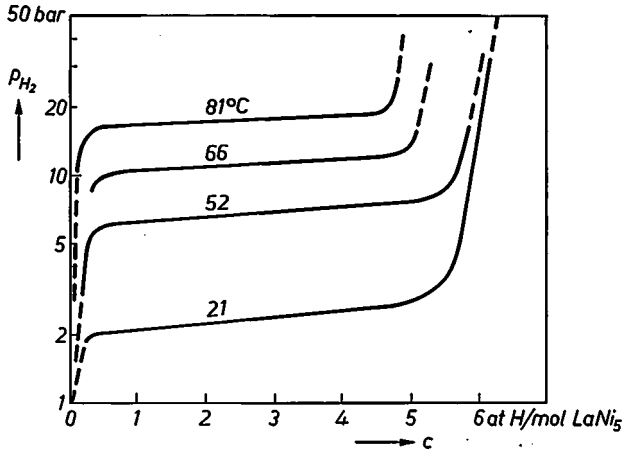


Fig. 15. The pressure of hydrogen p_{H2} in equilibrium with $LaNi_5$ as a function of the hydrogen concentration c at different temperatures. If for example the H_2 pressure is raised to greater than about 3 bars at 21 °C, hydrogen is absorbed until the composition $LaNi_5H_6$ is reached approximately. Conversely, the hydrogen thus absorbed is released again when the pressure is lowered to less than 2 bars.

The metal behaves like a 'hydrogen sponge'. During the formation of the hydride it does indeed swell, growing by 25% of its original volume, just like a sponge absorbing water.

An effect of this expansion is illustrated in *fig. 16*, where a small piece of freshly prepared metal exposed for the first time to hydrogen gas is observed to crumble spontaneously into small crystallites. The clean, active surface formed during the fracture and fragmentation has the effect of speeding up the reaction. The crumbling continues until a very fine powder is formed, with an average grain diameter of 5 μm and a very large and highly reactive surface that has a high sorption rate even down to well below 0 °C.

The potential applications of a substance possessing such properties are legion. As a *storage medium* for hydrogen it has advantages over steel gas cylinders or Dewar vessels for liquid hydrogen. The hydrogen density in the $LaNi_5H_{6.7}$ lattice is nearly twice as great as in liquid hydrogen. In addition, a tank containing $LaNi_5H_{6.7}$ does not need to have a particularly thick wall, since the pressure in it at room temperature is

only 2.5 bars. The heat of dissociation of the hydride (approximately 7.7 kcal/mol H_2) could be used in a *gas-refrigerating machine* (cryogenerator). By making use of the temperature dependence of the equilibrium pressure above the hydride it has proved possible to make a *thermocompressor* with no moving parts, intended for a hydrogen cryogenerator in the 20 K temperature region [19]. The hydride is also suitable for use as an electrode in *rechargeable batteries*. Unlike conventional electrode materials, this hydride retains metallic conduction during the chemical reactions in the battery. $LaNi_5$ is used as a *pressure buffer*, e.g. in combination with nickel-nickel-hydroxide batteries in space satellites. The use of $LaNi_5$ avoids the formation of excessive hydrogen pressures while these batteries are being charged by solar radiation.

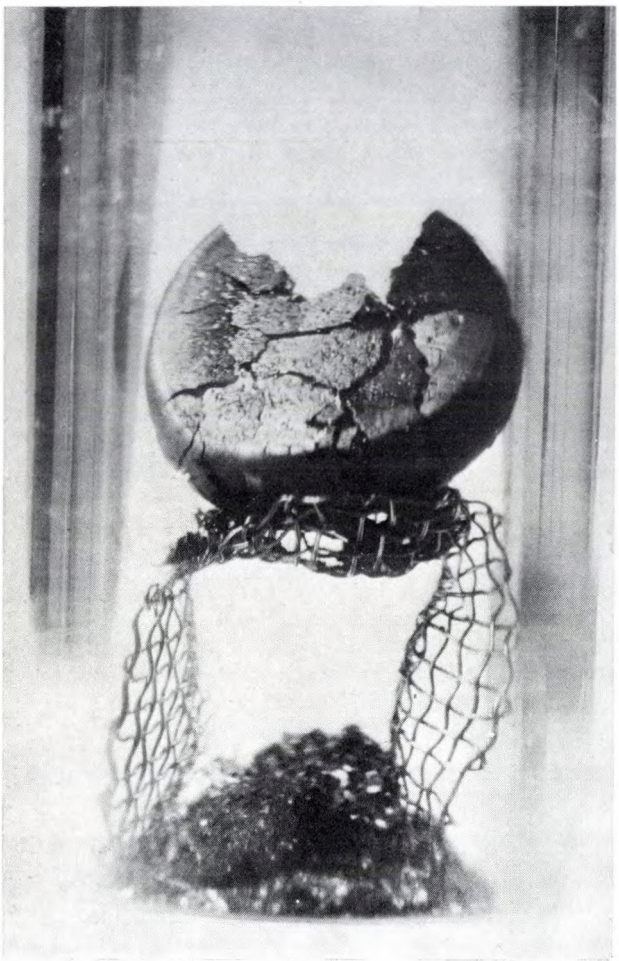


Fig. 16. The destructive effect of the spontaneous take-up of hydrogen by an $LaNi_5$ casting, demonstrated by letting this take place at high pressure on a piece of gauze in a thick-walled cylinder. The strongly expanding metal gradually crumbles into hydride grains that pile up underneath the piece of gauze.

Much attention has also been given to the possible use of $LaNi_5$ as a *hydrogenation catalyst*. The reaction between impurities and the lanthanum at the surface produces an oxide with the release of pure nickel, which is a highly active catalyst, in much the same way as the well known Raney nickel. Finally, an obvious application of $LaNi_5$ is its use as a *purification medium* for hydrogen, since hydrogen and its isotopes are the only gases that are able to penetrate into the $LaNi_5$ lattice.

As with other materials, the special requirements imposed by different applications can be met by $LaNi_5$ over a certain range. As an electrode material in batteries, for example, it is preferable to use $LaNi_5$ in which copper atoms have been introduced in a disordered way to occupy nickel sites. This material has better mechanical properties than pure $LaNi_5$. Of course, this also changes the equilibrium pressure and storage capacity of the material. Generally speaking, all kinds of properties can be changed 'to order' by the substitution of related elements for La or Ni.

[19] H. H. van Mal, Chemie-Ing.-Technik **45**, 80, 1973.

Summary. The background and results of 20 years of research on intermetallic compounds are described. It is shown how knowledge acquired in the investigation of getters has led to discoveries in entirely different fields. The results include the elucidation of the structure of various tri-aluminides of rare-earth metals, insight into the shape-memory effect of alloys such as NiTi, the development of material with a high $(BH)_{max}$ value for permanent magnets and the discovery of the capacity of $LaNi_5$ for absorbing hydrogen in large quantities and readily giving it up again.

Recent 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*
Mullard Research Laboratories, Redhill, Surrey, England *M*
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*
MBLE Laboratoire de Recherches, 2 avenue Van Becelaere, 1170 Brussels (Boitsfort), Belgium *B*
Philips Laboratories, 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A. (by contract with the North American Philips Corp.) *N*

Reprints of most of these publications will be available in the near future. Requests for reprints should be addressed to the respective laboratories (see the code letter) or to Philips Research Laboratories, Eindhoven, The Netherlands.

J. P. Beasley & D. G. Squire: An electron beam mask-maker.
IEEE Trans. **ED-22**, 376-384, 1975 (No. 7). *M*

A. J. van Bommel, J. E. Crombeen & A. van Tooren: LEED and Auger electron observations of the SiC(0001) surface.
Surface Sci. **48**, 463-472, 1975 (No. 2). *E*

M. Brouha & K. H. J. Buschow: Magnetic properties and pressure dependence of the Curie temperature of $LaCo_{5x}Cu_{5-5x}$.
J. appl. Phys. **46**, 1355-1358, 1975 (No. 3). *E*

T. M. Buck, Y.-S. Chen, G. H. Wheatley (all with Bell Laboratories, Murray Hill) **& W. F. van der Weg** (Philips Research Labs., Amsterdam Division): Energy spectra of 6-32 keV neutral and ionized Ar and He scattered from Au targets; ionized fractions as functions of energy.
Surface Sci. **47**, 244-255, 1975 (No. 1).

K. H. J. Buschow: Crystal structure and magnetic properties of $YFe_{2x}Al_{2-2x}$ compounds.
J. less-common Met. **40**, 361-363, 1975 (No. 3). *E*

K. H. J. Buschow, J. P. de Jong (University of Leiden), **H. W. Zandbergen** (Univ. of Leiden) **& B. van Laar** (Reactor Centrum Nederland, Petten): Magnetic properties of some light rare-earth compounds with CsCl structure.
J. appl. Phys. **46**, 1352-1354, 1975 (No. 3). *E*

R. Cadoret (U.E.R. Sciences, Aubière), **L. Hollan, J. B. Loyau** (C.N.R.S., Orléans), **M. Oberlin** (C.N.R.S., Orléans) **& A. Oberlin** (C.N.R.S., Orléans): GaAs growth by vapour phase transport, II. Interpretation of the growth of the {001} faces by the adsorption of gallium monochloride and arsenic molecules.
J. Crystal Growth **29**, 187-194, 1975 (No. 2). *L*

T. Claasen, W. F. G. Mecklenbräuker & J. B. H. Peek: Frequency domain criteria for the absence of zero-input limit cycles in nonlinear discrete-time systems, with applications to digital filters.
IEEE Trans. **CAS-22**, 232-239, 1975 (No. 3). *E*

C. D. Corbey: Simple TRAPATT circuit for 2nd-harmonic extraction.
Electronics Letters **11**, 283-284, 1975 (No. 13). *M*

J. P. Deschamps & A. Thayse: Representation of discrete functions.
Proc. 1975 Int. Symp. on Multiple-valued logic, Bloomington, pp. 99-111. *B*

W. F. Druyvesteyn: Thickness and magnetization variation in magnetic bubble materials.
J. appl. Phys. **46**, 1342-1348, 1975 (No. 3). *E*

H. C. van den Elzen: On the theory and the calculation of worst-case eye openings in data-transmission systems.
Philips Res. Repts. **30**, 385-435, 1975 (No. 6). *E*

E. Fabre, M. Mautref & A. Mircea: Trap saturation in silicon solar cells.
Appl. Phys. Letters **27**, 239-241, 1975 (No. 4). *L*

C. T. Foxon & B. A. Joyce: Interaction kinetics of As_4 and Ga on {100} GaAs surfaces using a modulated molecular beam technique.
Surface Sci. **50**, 434-450, 1975 (No. 2). *M*

P. J. van Gerwen, W. F. G. Mecklenbräuker, N. A. M. Verhoeckx, F. A. M. Snijders & H. A. van Essen: A new type of digital filter for data transmission.
IEEE Trans. **COM-23**, 222-234, 1975 (No. 2). *E*

J. P. Hazan, L. Jacomme & D. Rossier: Temporal behaviour of a localized index-gradient fibre using variable-angle injection.
Optics Comm. **14**, 368-373, 1975 (No. 3). *L*

H. Kalis: DDA: Schnelle digitale Rechenelemente.
Elektronik (München) **24**, 59-62, 1975 (No. 5). *H*

E. T. Keve: Observation of orthorhombic→tetragonal→cubic phase transitions in some PLZT ceramics.
Appl. Phys. Letters **26**, 659-661, 1975 (No. 12). *M*

H. Köstlin, R. Jost & W. Lems: Optical and electrical properties of doped In_2O_3 films.
Phys. Stat. sol. (a) **29**, 87-93, 1975 (No. 1). *A*

H. Kurz & E. Krätzig: Spectroscopic investigation of the photorefractive recording and erasure process in $LiNbO_3$:Fe.
Appl. Phys. Letters **26**, 635-637, 1975 (No. 11). *H*

J. B. Loyau, M. Oberlin, A. Oberlin (all with C.N.R.S., Orléans), **L. Hollan & R. Cadoret** (U.E.R. Sciences, Aubière): GaAs growth by vapour phase transport, I. Study of the effect of supersaturation and surface adsorption.
J. Crystal Growth **29**, 176-186, 1975 (No. 2). *L*

C. Michel: Effect of substrate microdefects due to thermal etching on the near-junction properties of GaP:Zn,O grown by liquid phase epitaxy.
J. Electrochem. Soc. **122**, 678-682, 1975 (No. 5). *N*

M. Redwood, R. B. Topolevsky (both with Queen Mary College, London), **R. F. Mitchell & J. S. Palfreeman:** Coupled-resonator acoustic-surface-wave filter.
Electronics Letters **11**, 253-254, 1975 (No. 12). *M*

J. M. Robertson, J. C. Verplanke, S. Wittekoek, P. F. Bongers, M. Jansen & A. Op den Buis: Non-stoichiometry and magneto-optic properties of magnetic bubble thin films of TmIG substituted with bismuth and gallium.
Appl. Phys. 6, 353-356, 1975 (No. 3). *E*

E. W. F. Roeder & H. J. C. Wilson: The system CaO-ThO_2-ZrO_2.
J. Amer. Ceramic Soc. **58**, 161-163, 1975 (No. 5/6). *A*

T. E. Rozzi & W. F. G. Mecklenbräuker: Wide-band network modeling of interacting inductive irises and steps.
IEEE Trans. **MTT-23**, 235-245, 1975 (No. 2). *E*

J. P. Scott: An electron image projector with automatic alignment.
IEEE Trans. **ED-22**, 409-413, 1975 (No. 7). *M*

P. J. Severin, F. C. Eversteyn & F. Huizinga: On the measurement of the thickness of epitaxial and polycrystalline silicon layers in the submicron range.
Phys. Stat. sol. (a) **28**, 205-209, 1975 (No. 1). *E*

M. Sintzoff & A. van Lamsweerde: Constructing correct and efficient concurrent programs.
Proc. Int. Conf. on Reliable software, Los Angeles 1975, pp. 319-326. *B*

G. E. Thomas, E. E. de Kluizenaar & M. Beerlage: Light emission from excited molecules sputtered from chemisorption layers on silicon.
Chem. Phys. **7**, 303-309, 1975 (No. 2). *E*

J. M. P. J. Verstegen (Philips Lighting Division, Eindhoven) **& J. L. Sommerdijk:** Mn^{2+} and Tl^{+} luminescence in β-aluminas.
J. Luminescence **10**, 31-38, 1975 (No. 1). *E*

H. Weiss, E. Klotz & R. Linde: Deconvolution systems for coded aperture images of three-dimensional x-ray objects.
Optics and Laser Technology **7**, 117-120, 1975 (No. 3). *H*

H. W. Werner, A. E. Morgan & H. A. M. de Grefte: Interpretation of secondary ion mass spectra by means of fingerprint spectra and secondary ion imaging.
Appl. Phys. **7**, 65-69, 1975 (No. 1). *E, H*

M. V. Whelan, L. A. Daverveld & J. G. de Groot: Resistive-insulated-gate arrays and their applications: an exploratory study.
Philips Res. Repts. **30**, 436-482, 1975 (No. 6). *E*

Contents of Philips Telecommunication Review **33**, No. 4, 1975:

D. W. Rollema: The Schelde radar chain (pp. 177-190).
Fifty years ago (p. 191).
P. A. Carol: Maritime traffic systems using shore-based radar (pp. 192-203).
R. van Halen & H. Schreur: A telephone traffic measuring system (pp. 204-210).
A. M. Giacometti & T. F. S. Hargreaves: A 140 Mb/s digital transmission system for coaxial cables (pp. 211-218).

Contents of Valvo Berichte **19**, No. 5, 1975:

SOAR — Sicherer Arbeitsbereich für Transistoren (pp. 171-203).
K. Kröner: Elektronische Programmwahlschaltung mit integrierten Schaltungen für Rundfunk- und Fernsehempfänger (Touch control) (pp. 204-212).
Technologie moderner Kondensatoren (pp. 213-222).

PHILIPS TECHNICAL REVIEW

VOLUME 36, 1976, No. 6

Single-mask bubble memory with rotating-field control

W. F. Druyvesteyn, F. A. Kuijpers, A. G. H. Verhulst and C. H. M. Witmer

Current developments indicate that bubble memories will have capacities exceeding 10^6 bits and access times shorter than 1 ms. It therefore seems likely that they will be of greatest use in applications where 'high-speed' core or semiconductor memories are not large enough and conventional 'large' memories (magnetic tape and large discs) are not fast enough. The competitors for such applications are the charge-coupled devices. In both types the bits are moved through shift registers; in bubble memories this is done with the aid of a rotating magnetic field. The shift-register operation is obtained with circuits produced by photolithographic methods. The advantage of bubble memories is that only a few masks are required. 'Single-mask' memories are obviously particularly important here. Bubble memories are generally controlled by current pulses, and in some cases the rotating field that propels the bubbles also has a control function. When rotating-field control is used, the sense of rotation is reversed at preset times. For single-mask memories it would greatly simplify matters if complete control could be obtained via the rotating field alone. This article shows that rotating-field control is in fact satisfactory on a laboratory scale.

Introduction

In bubble memories the information is carried by 'magnetic bubbles'. The medium for the bubbles is a thin layer of ferromagnetic material magnetized perpendicularly to the layer by an external magnetic field H_0. The bubbles form stable but extremely mobile cylindrical domains of opposite magnetic polarity in the layer; see *fig. 1*. Such domains are possible only in material that is magnetically anisotropic, with the 'easy' direction of magnetization perpendicular to the layer, and in a relatively small interval of values of H_0. In the present experimental memories the diameter of the bubbles and the thickness of the layer are a few microns.

Series of preferred positions for the bubbles are created in the layer, usually by means of permalloy strips. A bubble in one of these positions represents the binary '1' and the absence of a bubble the '0'. A frequently used pattern of strips is illustrated in *fig. 2*; the details are of the dimensions of a bubble. The preferred positions depend on the direction of an auxiliary magnetic field H_1 in the plane of the layer. In fig. 2 this field is in direction *1*; in this situation the 'positive' bubble-attracting poles *1* form the preferred positions. If field H_1 is now rotated in a counter-clockwise sense, the bubbles jump successively to the new nearest preferred positions *2*, *3*, *4*, *1* . . . and thus move from the bottom to the top. The permalloy pattern therefore forms a shift register for the information. If the 'rotating field' H_1 rotates in the clockwise sense, the information shifts downwards. This type of shift register — with additional elements of other shapes — is

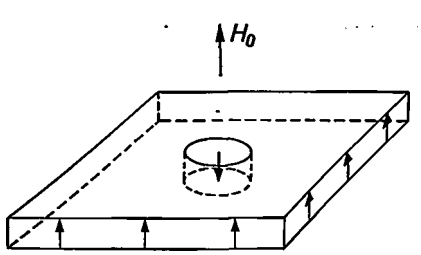


Fig. 1. The magnetic bubble, a cylindrical domain whose magnetization is in the opposite sense to that of the rest of the layer, which is magnetized in the direction of a constant external field H_0 perpendicular to the plane of the layer.

Dr Ir W. F. Druyvesteyn, Dr Ir F. A. Kuijpers, A. G. H. Verhulst and C. H. M. Witmer are with Philips Research Laboratories, Eindhoven.

employed in the majority of bubble memories developed to date, and we shall confine our discussions here to this type.

The way in which a computer memory is usually arranged is illustrated in *fig. 3*. The information is stored in a number of small loops L_1, L_2 . . . in which the bubbles move along tracks like those shown in fig. 2. These loops are connected to a large loop L_0, which is used for writing in, reading out and erasing information. To enable the arrangement to operate as a memory, there has to be a *bubble generator* (G), a *detector* (D), an *annihilator* (A) and also various *transfer gates* to enable bubbles to be switched from one track to the other at will.

These functions can be provided in a variety of ways. One method frequently used is to place a pattern of non-magnetic conductors underneath the permalloy pattern on the magnetic layer. Transfer gates, generators, etc. are then controlled by current pulses. An example is the transfer gate in *fig. 4*. The conductor is shown shaded. The rotation of the field H_1 is always counter-clockwise. If a bubble enters via track S_1, it normally leaves the transfer gate at S_2, via poles P and Q. If, however, a current pulse is sent through the conductor for a quarter-cycle of the rotating field starting when the bubble is at P, the bubble is subjected to a force to the left and directed into R. It then continues its path through S_3.

The production of such memories with a permalloy pattern and a conductor pattern makes great demands on technological skill. The patterns are applied to the magnetic layer by a photolithographic technique. Even in a simple experimental memory the permalloy pattern may easily consist of hundreds or even thousands of bars, tees and other elements measuring only several microns across, and application of such a pattern demands extreme care and accuracy (see for example fig. 12). The application, by means of two photomasks,

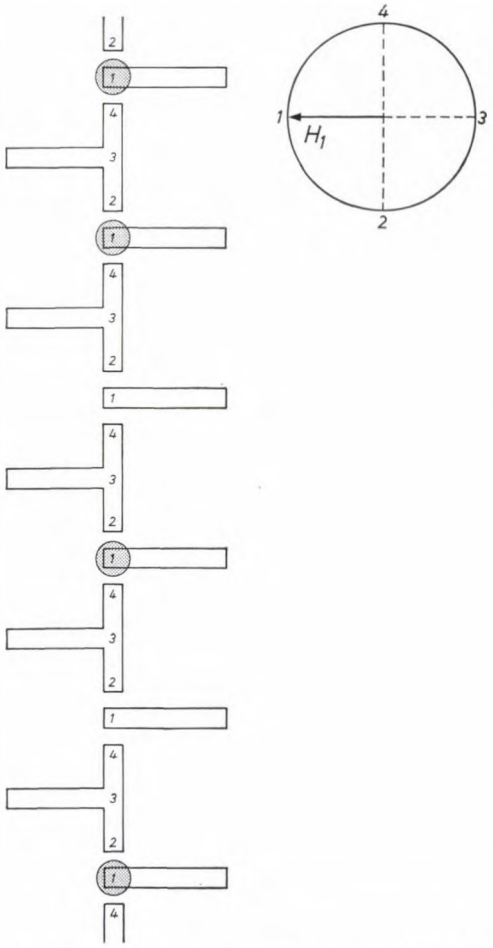


Fig. 2. Pattern of permalloy strips on a magnetic film, for guiding the bubbles. With the auxiliary field H_1 (in the plane of the film) in position *1*, the sites *1* form bubble-attracting poles; with H_1 in position *2*, sites *2* assume this function, and so on. If the field is rotated counter-clockwise (CCW rotation) the bubbles move upwards; clockwise (CW) rotation causes them to be displaced downwards.

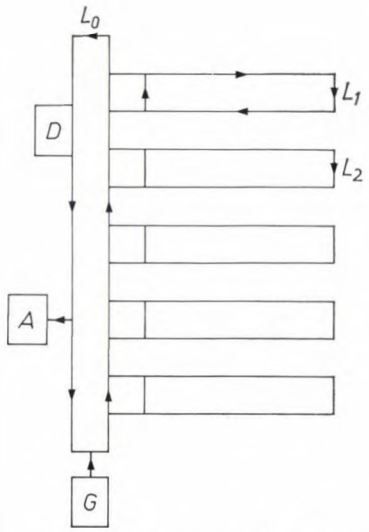


Fig. 3. Layout of a computer memory. The lines represent bubble tracks of the kind shown in fig. 2. L_1, L_2 . . . minor loops for the storage of information, L_0 major loop for write-in, read-out and erasure. G bubble generator, A annihilator, D detector. The transfer gates used to direct bubbles from one track to another are not separately indicated.

of two patterns accurately positioned in relation to each other adds an extra order of difficulty. 'Single-mask memories' — memories made with only one mask — are obviously at a great advantage here. Absence of the relative positioning requirement means not only that the production of memories for bubbles measuring about 5 μm across is a great deal simpler but also that it becomes more realistic to consider using submicron bubbles.

Single-mask memories can be controlled in two ways. Current pulses can also be used here, but now

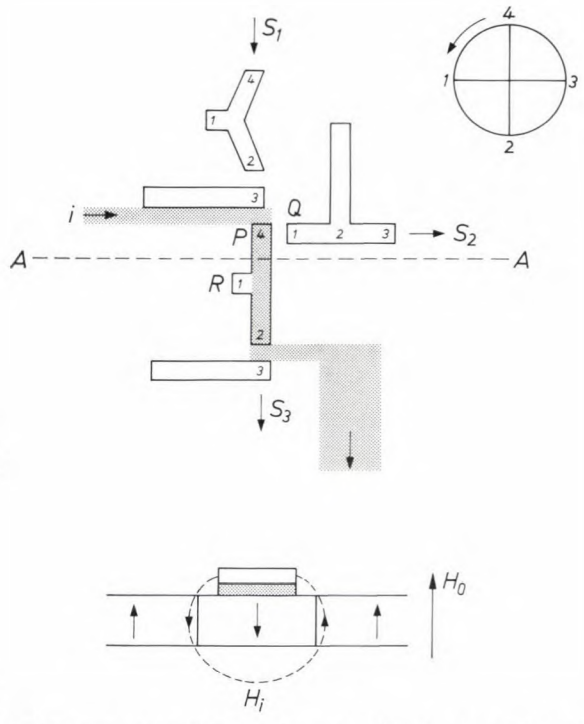


Fig. 4. Current-controlled transfer gate. H_1 always rotates in the CCW sense. *a*) The bubbles normally move from S_1 via poles P and Q on to S_2. At a particular current pulse i in the conductor (grey), a bubble at P is subjected to an extra force to the left, moves into R and passes on via S_3. *b*) Section A-A enlarged. The force to the left on the bubble arises because the field H_1 of the current through the conductor (grey, between the magnetic film and the permalloy strip) decreases the external field on the left and increases it on the right (see also fig. 9).

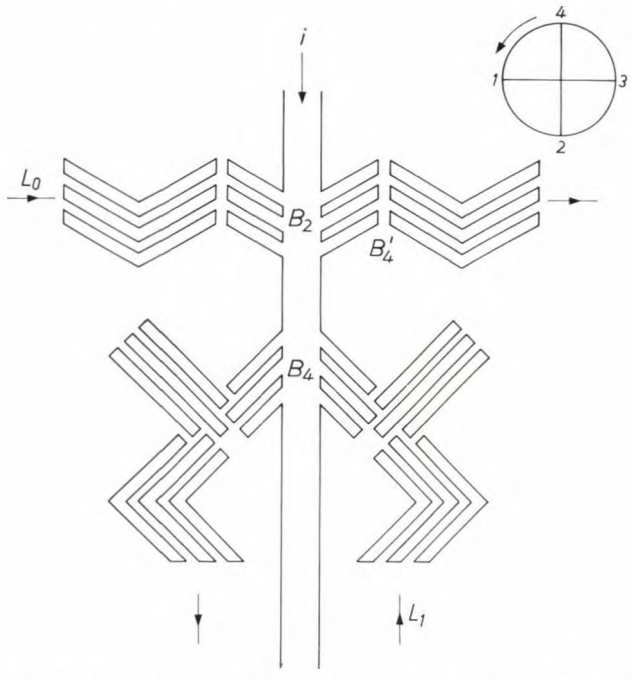


Fig. 5. Current-controlled transfer gate for a single-mask memory [1]. The bubble tracks in the major loop L_0 and the minor loop L_1 consist of chevrons. H_1 always rotates CCW. To transfer bubble B_2 (field position *2*) to site B_4 (field position *4*), i.e. from L_0 to L_1, a current pulse with a length of nearly one half-cycle is applied in the central permalloy strip. This holds the bubble in the transverse direction so that it moves towards B_4 in preference to B_4'. A current pulse between *4* and *2* can transfer a bubble from L_1 to L_0. By using different pulses the pattern can also act as a bubble splitter or 'replicator' which splits a bubble circulating in L_0 into two and diverts one into L_1.

they have to move through the permalloy. In this case the pattern has to incorporate not only pole-forming ends but also through-paths for the current. A configuration of this type [1] is illustrated in *fig. 5*; it can be used either as a transfer gate or as a bubble splitter. In practice the main drawback of this method is the rather large amount of heat developed because of the relatively high resistance of the permalloy.

The alternative is to use the rotating field for control. The direction of rotation of the field is reversed at certain times, which are specified in the control program.

Both methods have their advantages and disadvantages and a combination of the two will probably yield the best result eventually. In this article, however, we propose to show, with the aid of suitable patterns and the associated programs, that *rotating-field control on its own* is capable of operating a fairly complicated memory adequately, even though the field acts on all the bubbles simultaneously.

For completeness we will consider in more detail the structure of a 'memory chip', i.e. a chip several mm^2 in area containing one circuit of the type shown in fig. 3. A bubble memory normally contains many such chips, which in some types of memories are identical.

The structure of a memory chip

Probably the most successful media for bubbles are single-crystal films of certain types of iron garnet, epitaxially grown on a single-crystal substrate of gadolinium gallium garnet (GGG). The composition of the films which we use is approximately $Sm_{0.3}Y_{2.7}Ga_{1.2}Fe_{3.8}O_{12}$. The magnetic properties can be varied by modifying the composition. It is these properties, together with the thickness of the film and the external field, that determine the diameter of the bubble. A typical film is 5 μm thick and the diameter of the bubbles is 7 μm in an external field of about 50 Oe (4 kA/m). If the external field becomes too large, the bubble disappears; if it becomes too small, the bubble changes into an elongated domain (see *fig. 6*).

The speed of a memory is determined by the speed of rotation of the rotating field. The memories we shall discuss here continue to operate effectively at frequencies as high as 200 kHz. Using magnetic materials of different compositions it is not impossible to reach frequencies above 1 MHz.

[1] T. J. Nelson, AIP Conf. Proc. **18** (19th Ann. Conf. on Magnetism and magnetic materials, Boston 1973), Part 1, p. 95, 1974.

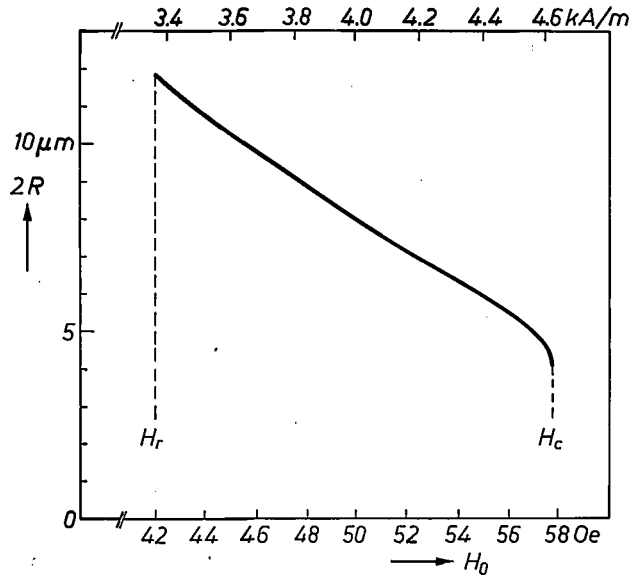


Fig. 6. The diameter $2R$ of a bubble as a function of the external field H_0. The bubble shrinks with increasing field, finally disappearing at the collapse field H_c. With decreasing field it changes at the run-out field H_r into a filamentary domain. The case illustrated above is that of a 5-μm film with a saturation polarization J_s of 15 mT (in the Gaussian system, a saturation magnetization $4\pi M_s$ of 150 gauss) in which the usual parameter l, indicating the characteristics of the film, is 1 μm. (In the Gaussian system $l = \sigma_w/4\pi M_s^2$, where σ_w is the energy per cm² of magnetic domain wall.)

The external field is usually generated by coils or permanent magnets; in general magnets are to be preferred because the contents of the memory will not be erased if the power fails. The rotating field is produced by crossed coils carrying currents with a relative phase difference of 90°. With rotating-field control it has to be possible to reverse the sense of rotation in each of the four field positions (see fig. 2) within a period that is very short in relation to the cycle of the rotating field. This has been successfully achieved at Philips Research Laboratories at a frequency of 100 kHz using current generators that can deliver currents of 3 A into coils of 20 μH [2].

We shall now explain in somewhat greater detail the behaviour of a bubble when the external field is varied.

Let us assume that the energy of a layer of infinite extent of thickness d, with no bubbles and subject to an external magnetic field H_0 directed upwards is zero. The generation of a bubble in this layer can mean a reduction in energy, because the demagnetizing energy E_d is negative, since the spins adjacent to the bubble produce a downward demagnetizing field at the bubble (see *fig. 7*), and a bubble in that field has a lower energy (E_d) than an upward magnetized area would have. E_d increases in absolute value with the radius R of the bubble (see *fig. 8*). Two loss terms have to be set against this. Because of its opposite magnetization the bubble has a positive field energy in the external field H_0; this is proportional to its volume, i.e. to R^2d. Moreover, the energy E_w of the domain wall is positive because the spins in the wall are at an angle to each other and are not in the preferred direction. This contribution is proportional to the wall area, i.e. to Rd.

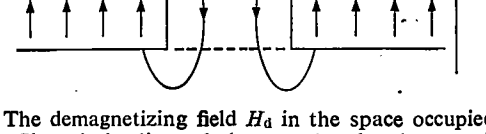


Fig. 7. The demagnetizing field H_d in the space occupied by a bubble. Since it is directed downwards, the demagnetization energy E_d of a bubble is negative.

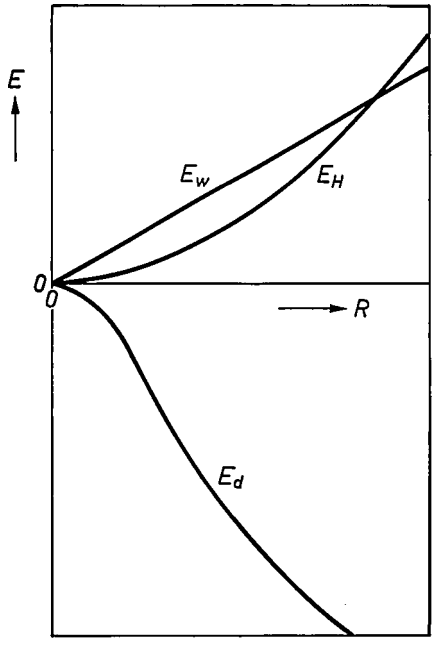


Fig. 8. Contributions to the energy of a bubble. E_d demagnetization energy, E_H field energy, E_w domain-wall energy.

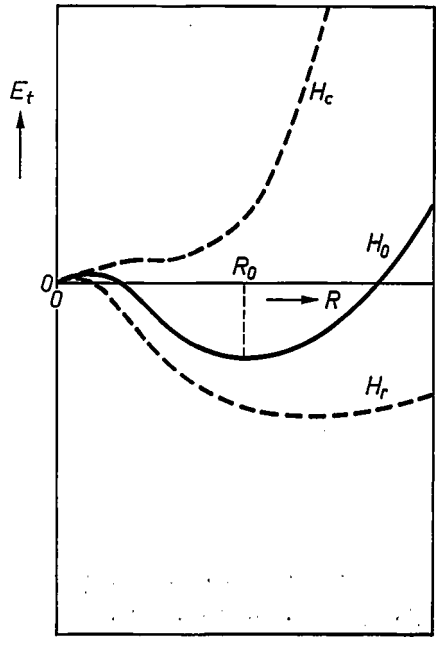


Fig. 9. Total energy E_t of a bubble in an external field H_0. The bubble is stable at a radius R_0 at which E_t is a minimum. Dashed lines: the total energy if the external field is equal to the collapse field H_c or the run-out field H_r ($H_r < H_0 < H_c$); in the first case E_t only just fails to have a minimum. For a decreasing field the minimum becomes lower; bubbles therefore move in the direction of decreasing field-strength if the field has a gradient.

The sum E_t of the contributions in fig. 8 as a function of R sometimes has a minimum (*fig. 9*) because, as R increases, the decreasing E_d is at first dominant, but later E_H becomes dominant. The value of R at this minimum is the radius assumed by the bubble. E_H increases with increasing field H_0; the minimum moves to the left and the bubble therefore becomes smaller. The minimum also becomes flatter; at the 'collapse field' H_c it disappears and so does the bubble (see fig. 6). For a decreasing field the radius increases until at field H_r — the 'run-out field' — the cylinder is no longer stable for elliptical deformations and changes to an elongated domain.

As the field decreases not only does the bubble become larger but the energy minimum becomes lower (fig. 9). In a spatially modulated external field the bubbles therefore move in the direction of decreasing field. The operation of the permalloy poles is based on this effect: a pole attracts if it lowers the external field locally. It can also be seen that bubbles have to repel one another: they increase the field in their vicinity.

Fig. 10 is a schematic diagram of a section through a memory chip. The diagram shows the substrate (*GGG*), the magnetic film (*IG*), the permalloy layer (*P*) and a gold layer (*Au*) for contacts, and also a dielectric intermediate layer (*Q*) and a layer implanted with ions (*I*) in the iron garnet. The dielectric layer, usually of SiO_2, improves the mobility of the bubbles. If it is not present the bubbles, which are themselves 'magnets', stick too easily to a permalloy pole-forming projection when this is nominally no longer a preferred position, like magnets to a soft-iron sheet. This adhesive effect decreases rapidly with increasing thickness of the intermediate layer.

Ion implantation is one of the methods used for suppressing 'hard' bubbles. These are unwanted bubbles with a highly deviant behaviour. They are stable over a much wider range of values of the external field and display a much greater variety in dimensions than ordinary bubbles. Moreover, they are not very mobile. These and other properties are the result of a complicated structure of the domain wall. We will not discuss this in greater detail here [2].

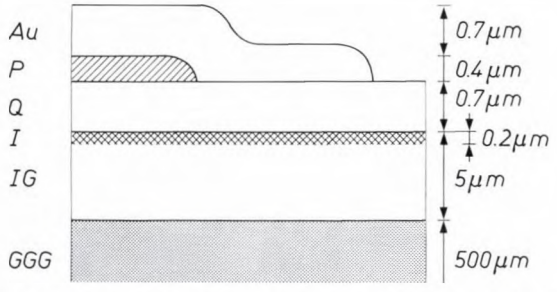


Fig. 10. Cross-section through a memory chip, schematic. *GGG* substrate. *IG* magnetic film of iron garnet. *Q* dielectric intermediate layer of quartz glass. *P* permalloy pattern. *Au* pattern of gold for contacts. A thin layer *I* of the iron garnet is implanted with ions to suppress 'hard' bubbles. A layer of titanium oxide (5 nm) between *Q* and *P* to improve the adhesion of the permalloy and a layer of molybdenum (0.1 μm) on *P* and *Q* to improve that of the gold are not shown.

A memory chip of the kind illustrated in fig. 10 is produced in the following way at Philips Research Laboratories. The iron-garnet film is deposited on the substrate epitaxially from the liquid phase [3] and bombarded by Ne^+ ions with an energy of 100 keV and at a density of 2×10^{14} ions per cm^2 ('ion implantation') [4]. Then the surface of the chip is thoroughly cleaned, and the dielectric layer, a 5-nm film of titanium oxide (not shown in fig. 10), and the permalloy film are deposited on it in turn by sputtering. The titanium oxide is applied to ensure good adhesion of the permalloy. The sputtering takes place in a 13.56-MHz argon discharge, using cathodes with coatings of quartz glass, titanium or a permalloy consisting of 80% by weight of nickel and 20% by weight of iron. In the case of the titanium oxide the atmosphere contains 3-4% of oxygen in addition to the argon; the film here is formed by 'reactive sputtering' of the titanium-coated cathode.

The pattern in the permalloy is created by a photolithographic technique in a manner similar to that adopted for ICs. A solution of positive photoresist is distributed very evenly over the surface and dried by rapidly rotating the chip with a few drops of solution on it (3 s at 10 000 revs/min and 30 s at 3000 revs/min). The resist is then exposed through a mask carrying the appropriate pattern. After processing, the pattern remains on the permalloy as photoresist.

The permalloy left uncovered has then to be removed by etching. In view of the accuracy required in the lateral direction — a fraction of a micron — we favour etching by sputtering ('ion etching') [5]. In this process the chip is attached to the cathode and bombarded by Ar^+ ions from an r.f. discharge. Since both the permalloy and the resist are etched, the resist film has to be so thick that it does not disappear before the permalloy. As a result of the ion bombardment the resist that is left is hardened and special resist removers have to be used. It can also be removed ultrasonically. However, the permalloy need only be bared where it will later have to make contact with the gold coating. Details of a permalloy pattern after this process can be seen in *fig. 11*.

To ensure better adhesion of the gold a 0.1-μm molybdenum layer is deposited by sputtering before the gold layer itself. Photolithographic methods are used to form the gold strips that connect the permalloy

[2] See W. F. Druyvesteyn, A. W. M. van den Enden, F. A. Kuijpers, E. de Niet and A. G. H. Verhulst, in: Solid state devices, 1974, Proc. Conf. Nottingham (Inst. Phys. Conf. Ser. No. 25), p. 37, 1975.

[3] For a description of this process see J. A. Pistorius, J. M. Robertson and W. T. Stacy, Philips tech. Rev. **35**, 1, 1975.

[4] This is done at the Institute for Nuclear Physics Research, Amsterdam.

[5] See for example H. Dimigen and H. Lüthje, Philips tech. Rev. **35**, 199, 1975.

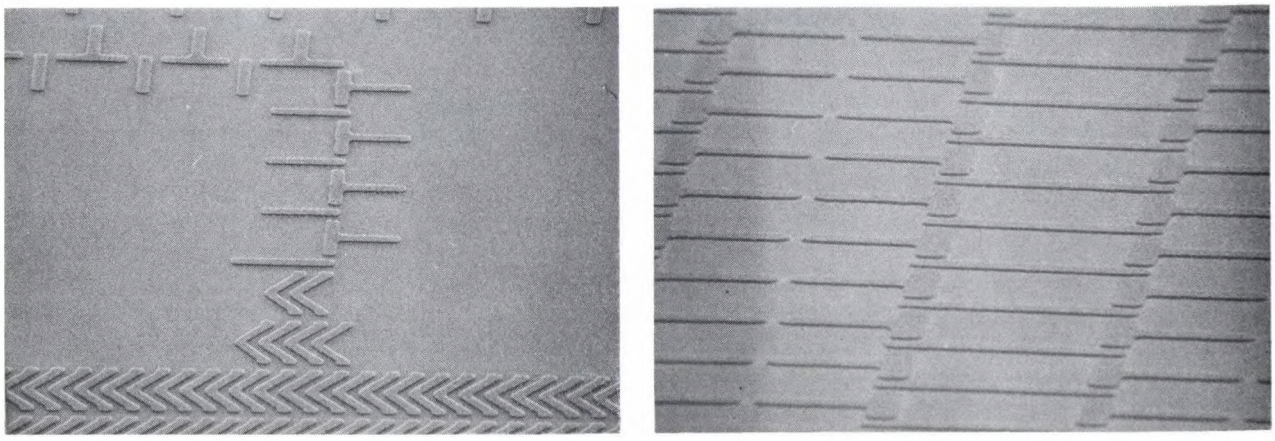


Fig. 11. Scanning-microscope photographs of permalloy patterns. The magnifications are 420× *(left)* and 830× *(right)*. The directions of viewing are at angles of 30° and 15° respectively to the surface.

detector to the detection equipment. Since these are broad strips and no great accuracy is necessary (see fig. 12), the disadvantage of the two-mask technique is not found here. The gold is etched by a wet-chemical process. The entire surface is then covered with a protective coating. A number of circuits are made simultaneously on a single chip; the circuits are separated by scribing with a laser beam and breaking.

Single-mask memories with rotating-field control

We shall now discuss the method of control with a rotating field by reference to an example: the memory circuit [6] in *fig. 12*. It will be obvious from its very modest dimensions that this memory is not suitable for immediate application but is intended merely to demonstrate the possibilities of rotating-field control. The circuit follows the scheme shown in fig. 3: a major loop L_0 for write-in and read-out, six minor loops $L_1, \ldots, L_6$ for the actual storage, a bubble generator G, an annihilator A_1 and a detector D. We shall first consider briefly the generator, the annihilator and the transfer gates and then the organization of the circuit as a whole. Then we shall look at detection more closely.

The bubble generator is illustrated in *fig. 13*, where a series of photographs also show it in action [7]. A master domain circulates permanently under the square permalloy area. When the field rotates from position *3* to position *4*, the projection A forms a strong negative pole that assists the movement from B to C. The negative pole formed at D in position *1* cuts through the domain that has meanwhile become elongated. We have found that a small projection E (not present in the generator illustrated in the photographs) assists this division. When the rotation is counter-clockwise, the generator produces an uninterrupted stream of bubbles. We will consider the generation of information from this stream later.

When the sense of rotation is reversed a bubble annihilator must not start to act as a bubble generator. The annihilators shown in fig. 12 — A_1, A_2 and A_3 — are not ideal in this respect and we shall limit our discussion here to the improved version shown in *fig. 14*. A bubble that has arrived at A in position *1* of the rotating field (fig. 14*b*) is made to merge with the

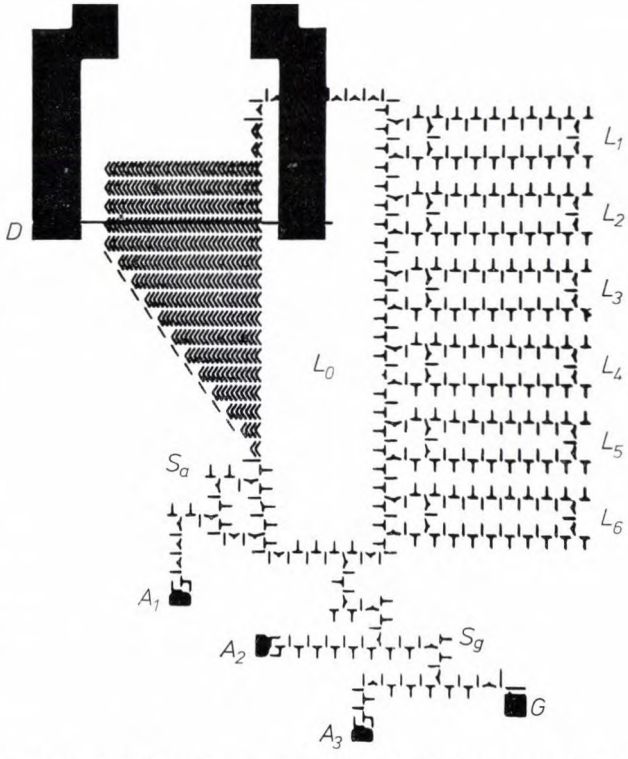


Fig. 12. Experimental memory circuit arranged in the layout of fig. 3. L_0 major loop. $L_1, L_2 \ldots L_6$ minor loops. G generator. A_1 annihilator. S_g generator lock (see p. 157). S_a annihilator lock. Annihilators A_2 and A_3 form part of the generator lock. Magnification 100×.

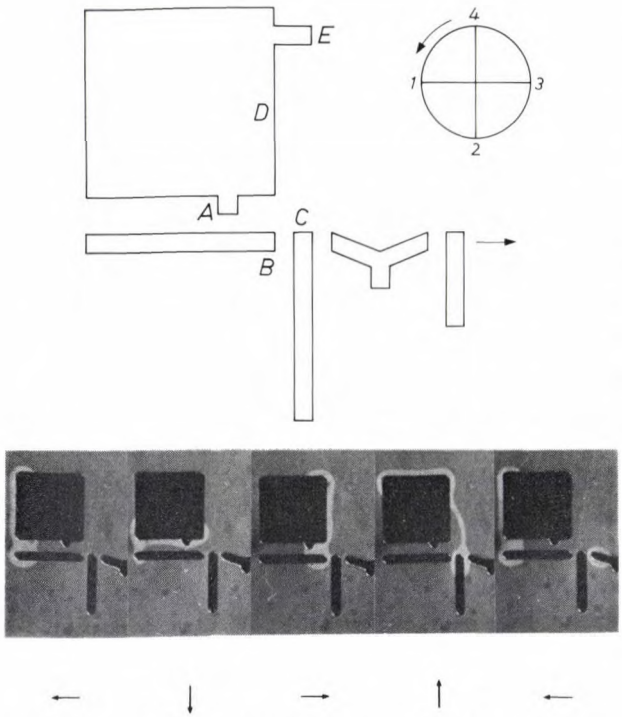


Fig. 13. *Above:* pattern of a bubble generator [7]. *Below:* the generator in action.

large domain under the chip despite the mutual repulsion of domains if the rotation is counter-clockwise. This process is largely due to the strong repulsion by D in position *2*. To increase this repulsion and reduce the attraction by E, the vertical bar of the tee is lengthened. When the rotation is clockwise (fig. 14*c*), the large domain under F in position *3* does not tend to cross to A on transition to position *2*, but is driven into pocket

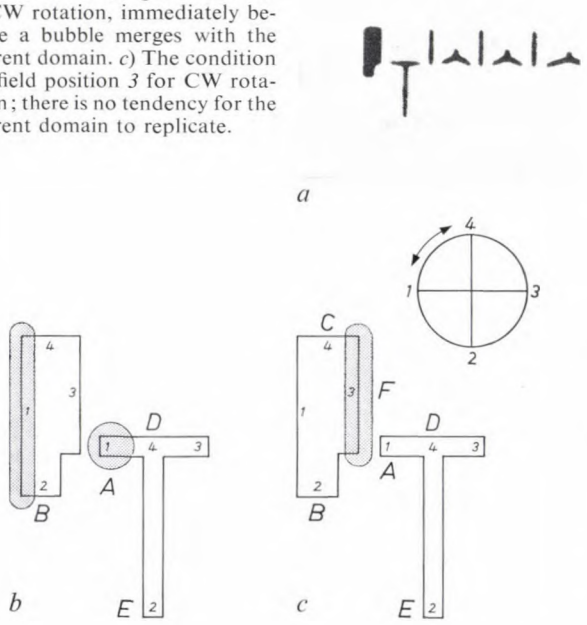


Fig. 14. *a*) Annihilator. *b*) The condition in field position *1* for CCW rotation, immediately before a bubble merges with the parent domain. *c*) The condition in field position *3* for CW rotation; there is no tendency for the parent domain to replicate.

B by the attraction of B *plus* the repulsion of C and D.

The principle of the transfer gate with field control is illustrated in *fig. 15* [8]. The dots represent the preferred positions. The normal direction of rotation is counter-clockwise; the bubbles then move from track S_1 to track S_2. For a bubble at the switch point B (field position *1*) the nearby position C is designed to be more attractive than A if the field rotates clockwise to *4*. Similarly, F has a greater attraction than D for a bubble at E when the field rotates from *2* to *3* (counter-clockwise). Consequently, a bubble at B in field position *1* ends up at G after three quarter-cycles of clockwise rotation, followed by three quarter-cycles of counter-clockwise rotation. We denote this program by $(1), -3, +3$, where the minus sign refers to clockwise (CW) rotation, and the plus sign to counter-clockwise (CCW) rotation. The number in brackets refers to the initial field position. The net number of cycles in this program is zero; consequently, all the other bubbles are back in their starting positions (*1*) when it ends.

Fig. 16 shows an application of this principle. The number of quarter-cycles between switch points B and E is not critical, and there are seven in fig. 16. The

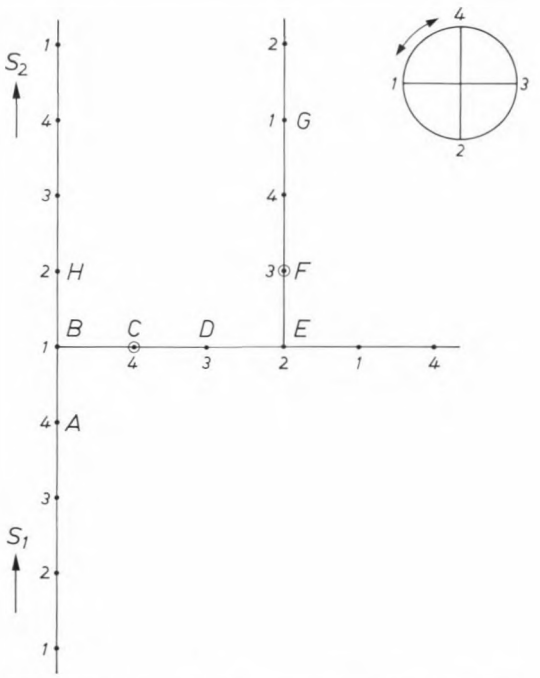


Fig. 15. Transfer gate with field control, schematic [8]. The successive poles are indicated by dots. In field position *4*, C exerts more attraction than A on a bubble in B, and in field position *3*, F exerts more attraction than D on a bubble in E. Thus the program 'from field position *1*, three CW quarter-cycles followed by three CCW quarter-cycles', causes a bubble with B as its starting point to move to G. After the program all other bubbles are back in their starting positions (*1*).

[6] W. F. Druyvesteyn, F. A. Kuijpers, E. de Niet and A. W. M. van den Enden, IEEE Trans. **MAG-10**, 757, 1974.
[7] R. F. Fischer, IEEE Trans. **MAG-7**, 741, 1971.
[8] P. C. Michaelis and I. Danylchuk, IEEE Trans. **MAG-7**, 737, 1971.

associated program is (*1*), —7, +7. The asymmetrical Y-elements are essential here; these give the preference for *C* and *F* to *A* and *D* discussed above.

Fig. 17 is a schematic diagram of the connection of a minor loop to the major loop. The arrows indicate the direction of movement for CCW rotation; the *1* positions are indicated by large dots. The program (*1*), —7, +7 takes a bubble that started off at *A* to *B* and thus conveys a bit from L_0 to L_1. Similarly, the program (*4*), —7, +7 takes a bit from *C* to *D*, i.e. from L_1 to L_0. The first program effects a write-in operation, the second a read-out operation. An essential requirement for the application is the possibility, created by the difference between the starting field positions, of executing the two programs independently of each other. While it is true that when the operation (*4*), —7, +7 is performed a bubble at *E* is sent to *F* after the first —7 and not to *G* where it came from, the fact that it does not reach the second switch point means that it is returned to *E* after the next +7.

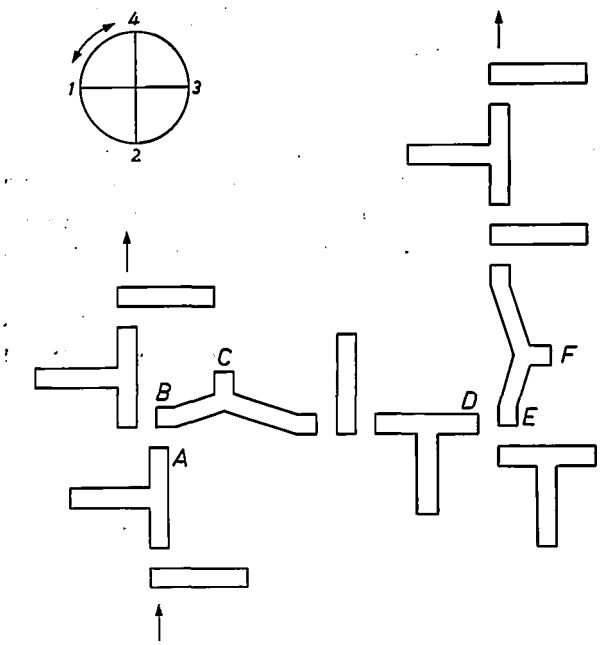


Fig. 16. The principle of fig. 15 in practice. The asymmetrical Y-elements create the required preference for *C* and *F* rather than *A* and *D*. The distance between the switch points *B* and *E* is seven quarter-cycles here; the switching program is (*1*), —7, +7.

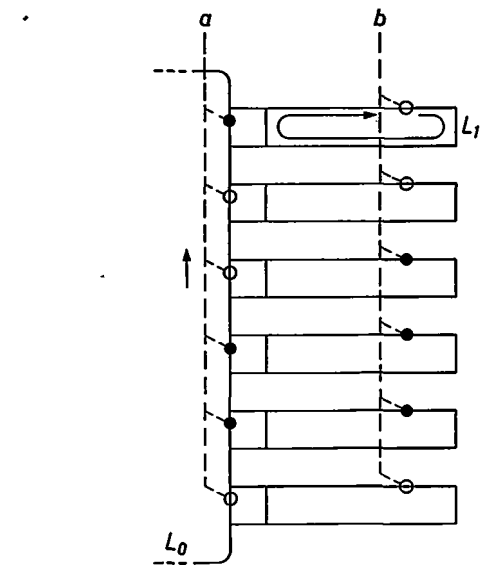


Fig. 18. Organization of the information. The distance between two bits of a word in L_0 is the distance between the inputs or outputs of the minor loops (four cycles); the number of bits is equal to the number of minor loops (six). Organized in this way, a word can be written in or read out as a single unit. With CCW rotation the word *a* (100110) circulates in L_0, the word *b* (001110) in the minor loops.

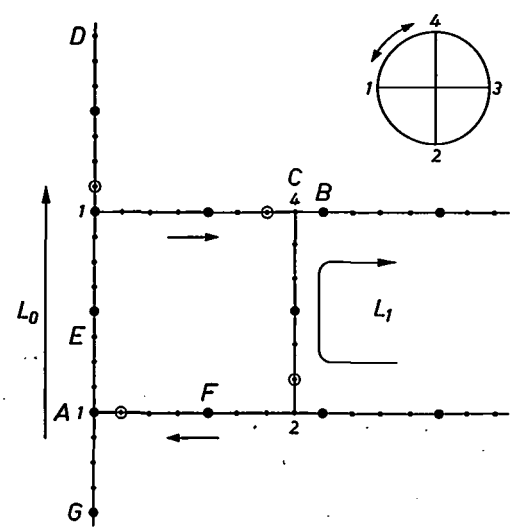


Fig. 17. Connection of a minor loop to the major loop, schematic. Sites *1* are indicated by large dots, the direction of movement for CCW rotation by arrows. The write-in program (*1*), —7, +7 moves a bubble from *A* to *B* and the read-out program (*4*), —7, +7 moves one from *C* to *D*.

The six minor loops in fig. 12 are all connected to L_0 in the same way; bits in corresponding positions are therefore written in or read out simultaneously. The interval between two successive inputs or outputs is four cycles. The information must therefore be organized in words of six bits at four-cycle spacings in L_0. In *fig. 18* the word *a* circulates in L_0, the word *b* in the six minor loops, with one bit in each loop. In the circuit shown in fig. 12 nineteen such words can be stored in the minor loops.

The information is generated with a bubble generator *G* and a transfer gate (*fig. 19*). The generator *G* sends out an uninterrupted stream of bubbles. The program (*2*), —7, +7, which, since the starting position (*2*) has not yet been used, is again independent of the two previous programs, takes a bubble from *P* to *Q* on the output track *S*. Unused bubbles disappear at *A*. Let the bubble at *Q* be the first bit ('1') of a word. The second bit must appear four cycles later: +16. For a '0' nothing more need be done; for a '1' we again have

−7, +7. In this way the word 110100, for example, is generated by (2), −7, +7, +16, −7, +7, +16, +16, −7, +7, +16, +16 = (2), −7, +23, −7, +39, −7, +39. The last bit has then arrived at Q. The net total (80) indicates the distance between the first bit and the last (5×4 cycles). Continued CCW rotation takes the word via a 'confluence transfer gate' into L_0; it can be stored when the first bit has reached the input to L_1.

A word that has to be read and put back into store is first read out with (4), −7, +7. It is read by the detector in the major loop. After 57 cycles, counting from the read-out, the word again appears at the inputs to the minor loops. The information stored in these loops has then circulated exactly three times, so that when the word is written in again it just 'drops into' the hole that has been left open. If the word is no longer required, the bits are conveyed one by one, via a transfer gate operated from the as yet unused field position 3, to the annihilator A_1.

Generator locks

To obtain the highest storage capacity it is desirable to operate various circuits simultaneously with a single set of field coils. If control is by the rotating field alone, the problem arises of how to process the information in one circuit independently of that in the other circuit. We shall now examine a simple example to show how 'generator locks' allow independent information to be generated in two circuits.

The generator locks for the two circuits *a* and *b* are represented schematically in *fig. 20*. Between the generator G and the output track S there are four switch points; the 'retrograde' (vertical) sections are seven quarter-cycles long; in between, there is a seven-quarter-cycle section in *a* and an eleven-quarter-cycle section in *b*. It can easily be seen from fig. 20 that the program P_a, (2), −7, +7, −7, +11 creates a bubble on the output track in circuit *a* while in circuit *b*, with continued CCW rotation, it only drives a bubble into the second annihilator. The opposite applies to program P_b, (2), −7, +11, −7, +7. In both cases all the other bubbles are advanced by one complete cycle.

To annihilate information it has also to be possible to select between the circuits, and annihilator locks can be constructed similarly for this purpose. No locks are required for writing into and reading out from the minor loops: each circuit has its own detector and it does not matter if words not read out continue to circulate in the major loop.

The more circuits there are from which to select, the longer the locks of the type in fig. 20 become, and hence the write-in time for new information. In a practical case, that of a 1.25-Mbit memory consisting of 64 chips, we have however calculated that the locks

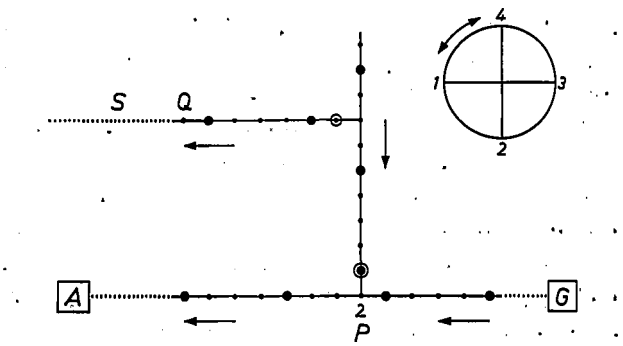


Fig. 19. Generation of information. By means of the transfer gate (program (2), −7, +7) any bubble generated by G can be directed as desired either on to the output track S or not. Unused bubbles disappear in the annihilator A.

increase the mean write-in time by a factor of only 1.7 and the mean annihilation time by a factor of 1.5 while the mean read-out time remains practically unchanged [6]. If the increase is too great, better locks can certainly be designed [2].

One of the practical disadvantages of circuits with locks is that they are all different and each will therefore have to be made with a separate photomask.

Detection of bubbles

Bubbles are in themselves not difficult to perceive. One of the attractive aspects of experimenting with bubbles is that, because of the Faraday effect and the

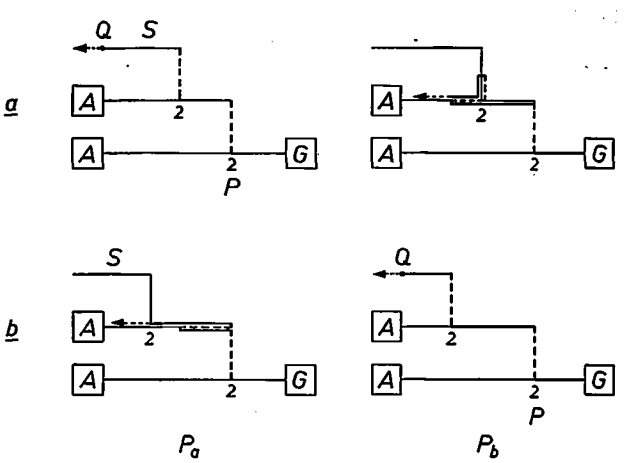


Fig. 20. Generator locks. A number of circuits are operated by the same rotating field and the locks are used to make one of these circuits at a time accessible for the bubbles generated. Each lock has its own 'key' program, which does not open the other locks. The generator locks for two circuits, *a* and *b*, are shown, with the tracks (thick) that a bubble passes along in the operation of the programs P_a and P_b:

$$P_a : (2), -7, +7, -7, +11$$

$$P_b : (2), -7, +11, -7, +7$$

The sections of track travelled with CW rotation are shown dashed. The program P_a takes a bubble to the output track S in *a*, but not in *b*; the converse is true for P_b. Thin dashed lines: track with continued CCW rotation.

transparency of garnets, they can be made visible with a polarizing microscope (see fig. 13). In a memory, however, 10^5 or 10^6 bubbles per second have to be converted one by one into a signal that can be processed electronically. The Faraday effect can also be used for this [9], but in most cases it is the magnetoresistance effect — i.e. the effect of the angle between the current and magnetization on the resistance — in ferromagnetic conductive materials, particularly permalloy, that is used.

Detectors based on this effect, including the one illustrated in fig. 12, consist of a permalloy strip that carries a current. A passing bubble affects the direction of magnetization and hence the voltage across the conductor. Its effect, however, is small. The greatest relative change in resistance — obtained when the saturation magnetization is rotated through 90° — is only 3% in permalloy and only a very small fraction of this is available in the actual application.

In the detectors that have been developed so far, two methods have been adopted to obtain sufficiently high sensitivity: the use of permalloy strips much thinner than the bubble tracks, and elongation of the bubble into a filamentary domain before detection. We considered thin detectors unsuitable since they have to be laid down as a separate coating close to or over the bubble tracks in an extra lithographic process and are therefore incompatible with a single-mask technology.

There are two reasons why thin detector strips are more sensitive than thick ones. First, the demagnetization factor in the plane vertical to the direction of flow is smaller, so that for a given change in the external field there is a greater change in the magnetization. Secondly, a thin detector gives a high voltage for a given measuring current. This current must not be too high because of its effect on bubble movement and the danger of burn-out.

The detector in fig. 12 is a somewhat modified version of the 'thick' detector developed by A. H. Bobeck [10] (see *fig. 21*). It forms an integral part of the bubble-track pattern and is therefore made at the same time. In it the bubbles are elongated considerably in a series of columns of 'chevrons'. (Chevrons are employed in numerous ways in bubble circuits; see fig. 5, for example.) Beneath a pole of one chevron in a column a bubble is not stable; on its arrival there it extends beneath the entire column. The actual detector is a column that can also function as a conductor for the measuring current, since it has connections between the chevrons.

The main considerations in designing a detection column are, on the one hand, to ensure the correct propagation characteristics for the chevrons and, on the other, to obtain a good detection signal. The detector column of fig. 12 differs markedly from that of

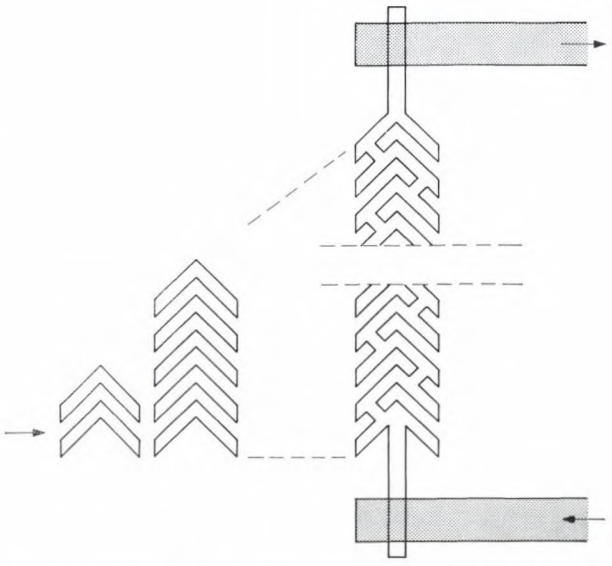


Fig. 21. Detector for single-mask memories [10]. By means of a series of columns of chevrons a bubble is elongated to the length of the actual detector. The figure shows a column carrying a measuring current, which is possible because the chevrons are interconnected. With constant measuring current the voltage variation occurring in each cycle as a result of the magnetoresistance effect depends on whether a bubble passes or not. The gold contacts are shaded. Elongation is necessary to obtain a usable detection signal. With a column consisting of 314 chevrons a 3-mV signal was obtained in the earliest detectors of this type with a measuring current of 5 mA for bubbles 6 μm across.

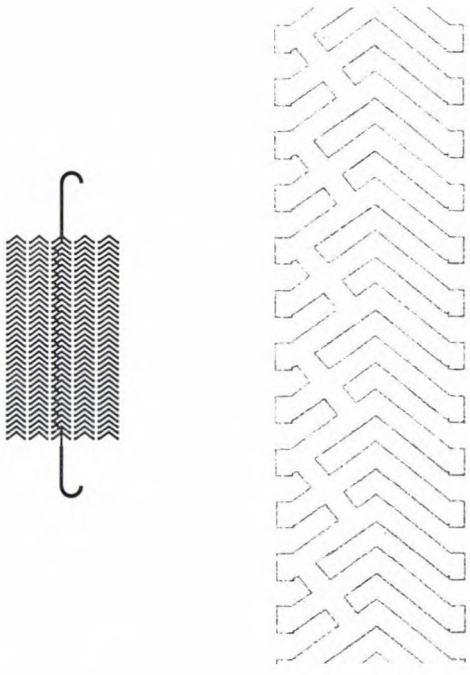


Fig. 22. A recent detector design with which we have obtained good results. *Left:* enlargement of the mask; *right:* detail of the pattern.

[9] See for example G. S. Almasi, IEEE Trans. **MAG-9**, 663, 1973.

[10] A. H. Bobeck, I. Danylchuk, F. C. Rossol and W. Strauss, IEEE Trans. **MAG-9**, 474, 1973.

fig. 21 in that the connections lie nearly in a single straight line. We were guided here by the consideration that a narrow resistance path should improve the signal-to-noise ratio and the discrimination between bubbles in close succession. *Fig. 22* shows a new design that gives very good results.

Finally, we ought to say something about the need to include the detector in a bridge with an identical dummy detector. The reason for this is that the rotating field itself is the largest source of variations in the magnetization and hence in the voltage across the detector; the difference between 'bubble' and 'no bubble' is relatively small. The variations occurring when no bubbles are passing are eliminated in the bridge.

The design of detectors and other elements is still largely empirical; the best design has to be chosen from perhaps 10 or 20 related versions based upon one particular idea. 'Calculation' of a detector is still impossible for various reasons, the main one being that the magnetic-domain structure in the permalloy is a vital factor, since the permalloy elements divide themselves into only a few domains. Magnetization in an element occurs as a result of wall displacement (*fig. 23*). These processes in a detector lead to considerable departures from the 'normal' magnetoresistance effect. In the 'normal' case only the mean magnetization M need be taken into account and the change in magnetization is proportional to $\sin^2 \alpha$, where α is the angle between M and the direction of the current flow.

* * *

As we pointed out earlier, the best possibilities for single-mask memories seem to be offered by a combination of current and field control. A combination that strikes us as particularly favourable is one in which only the generator is current-controlled. The information would then preferably have to be organized differently, not with a word stored in one chip (see fig. 18), but with the bits of one word being located at corresponding places in successive chips. The word length is then the number of chips. This system has various advantages. Limitation of current control to the generator keeps heat generation at very low levels. The chips can also be identical and therefore made with the same mask, generator locks being made unnecessary by the independence of the generators, while annihilator locks would become unnecessary because every word is annihilated or preserved in its entirety.

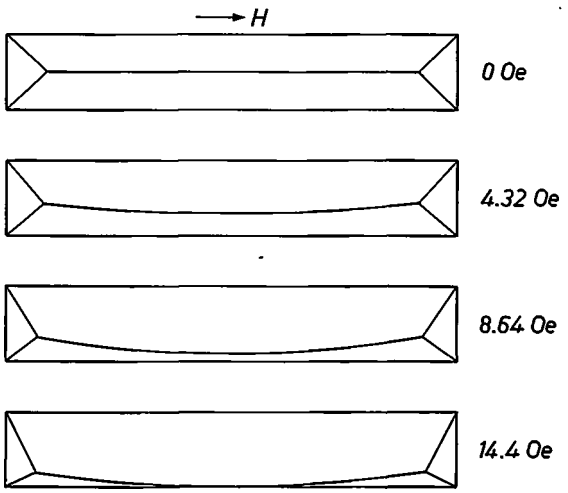


Fig. 23. Magnetization of a bar element, by domain-wall shift. These are tracings from photographs of a permalloy strip, 300 μm long, in which the domain walls were made visible by Bitter's method.

Summary. In magnetic-bubble memories the bubbles in a magnetic film are moved along tracks consisting of bars, tees or other shapes of permalloy elements on the film by means of a rotating field. The article describes a fairly complicated memory that can be completely controlled by the rotating field alone. In this it differs from memories that are wholly or partly controlled by pulses of current, the current flowing either in the permalloy or in a separate pattern of conductors. In the first case the considerable heat generated is a great disadvantage while in the second the technology is greatly complicated by the need to keep two patterns in register. Rotating-field control consists in reversing the sense of rotation at the appropriate moments. A basic feature of the method is the transfer gate, enabling a single bubble to be directed on to a different track while other bubbles are not affected. The bubble detector, operating by virtue of the magnetoresistance effect, is integrated in the permalloy pattern; to improve sensitivity a bubble is elongated before detection. Like the gold for the contacts and an intermediate film of silicon dioxide, the permalloy is applied to the magnetic film by sputtering. The same process is also used for etching the permalloy to the desired pattern. The tolerances for the gold strips are so large that the principal objection to multimask technology does not apply.



The 'Ergonomic', an ergonomically designed lathe, the result of cooperation with N.V. Hembrug. The operator can work comfortably at this lathe either seated or standing.

An ergonomic lathe

For about a year a rather unusual lathe has been in use in one of the workshops of the Philips Engineering Works in Eindhoven. The lathe is different from the others around it. First, the lathe bed is not horizontal but tilted towards the operator, and secondly the bed is located at a greater height from the floor. The design of this lathe is the result of the work of a group of specialists in ergonomics, industrial health, efficiency and construction and appraisal of machine tools.

This work was started because it had been established that lathe operators were particularly liable to lower-back pain. The main cause of these difficulties would appear to be the fact that lathe operators have to stand during the whole of the working day, with the back bent and the shoulders twisted in relation to the lower part of the body (see photo below on the left).

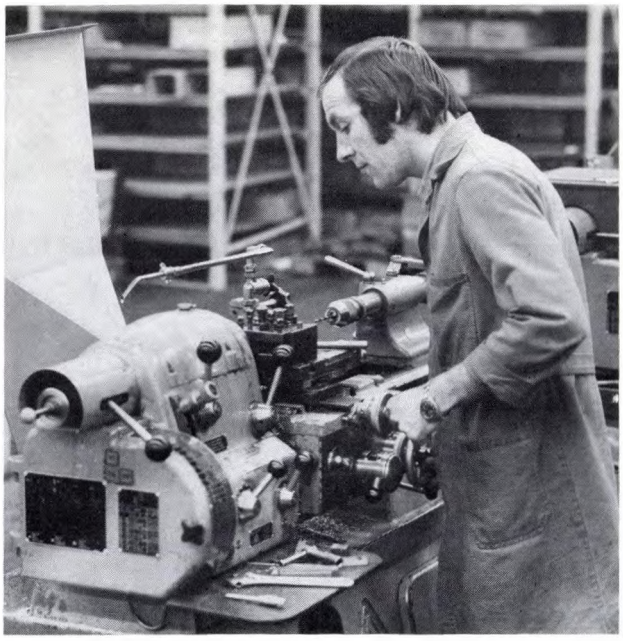
Starting from this point a number of requirements were formulated that an improved lathe ought to satisfy. Next, on the basis of these requirements a sheet-steel mock-up of a small lathe was made, which included the most important control units for such a lathe. The main features of this mock-up were the tilted bed mentioned earlier and a simple adjustment for the height of the lathe. This enabled the operator to have a good view of his work, while working in an easy, upright position. Knee room was also provided under the lathe, so that the operator could work just as well seated as standing. A large number of lathe operators were then asked to 'work' with the mock-up. In this way the design was further improved and the best positions for handwheels, operating levers and switches could be found.

At a later stage attention was paid to the graduated scales on the handwheels for the various tool slides. A graduation system was chosen consisting of fairly wide, matt-black stripes on a matt background. With this pattern the scales can be read off easily in the normal working position and annoying reflections from the lighting, etc., are avoided.

The experience gained with this mock-up has been put to use in a lathe constructed by N.V. Hembrug [*], in close cooperation with the designers. A prototype of this lathe, which has now been put on the market by N.V. Hembrug under the name 'Ergonomic', was shown at an exhibition of machine tools in Paris in June 1975.

A large number of lathe operators from various Philips workshops have worked with the machine shown in the photograph. This was done both to make the new lathe as widely known as possible inside the company, and also to obtain data for further improvements in the design.

[*] N.V. Hembrug, Machine-tool manufacturers, Zaandam, The Netherlands. The firm was formerly known as 'AI'.





Computer-aided design

P. Blume

In designing mechanical parts a great deal of the design time is taken up in making the drawings and updating them as required. Now that the cost of computer time is steadily decreasing, it is worth while investigating whether automation of these activities is a possibility and an economic proposition. Work has been in progress at the Philips Laboratories in Hamburg since 1973, in cooperation with colleagues at Philips Data Systems, Eiserfeld, on an integrated computer system in which parts are completely detailed in a dialogue between the designer and a computer via an 'interactive display'. The system also provides the programming of the numerically controlled manufacture of the part and delivers punched control tapes for these machines. For certain types of parts, such as punched metal parts, the tools required for the production process can also be designed in this way. The geometry of a workpiece can be stored in a data bank; the system employs an existing method developed for commercial data processing.

Introduction

Since the early sixties steadily increasing use has been made of computers in the design of mechanical devices and in production-planning departments — 'computer-aided design' and 'computer-aided manufacture', abbreviated to CAD/CAM. The objective is to make the computer do the routine work that people find so uninspiring. There is also the hope that the use of a computer will enable the design to be produced more rapidly and less expensively.

Before taking a closer look at the problems, let us first give a definition of the term 'mechanical design'. This term includes the assembly of a number of suitable components to form a complete machine as well as the separate geometrical specification of the individual parts. Both effectiveness and cost have to be taken into account, by choosing appropriate materials, making provision for strength requirements, following standards recommendations, and by keeping manufacturing as simple and inexpensive as possible [1]. The result of design is 'a large variety of technical data in forms such as drawings of parts and sub-assemblies, parts lists, adjustment instructions, manufacturing instructions, etc., enabling other people to manufacture the product' [2].

Stages in the design process

In engineering design the process of 'thinking out' a product — a process associated with the gradual taking shape of an idea that is originally undefined — is divided into several stages (see *fig. 1*) [3]. The first is the *concept stage* in which the original assignment, usually a list of requirements, is taken as the starting point in a search for design ideas that might offer a solution to the problem. The most suitable of these ideas is selected and is sketched by hand, not necessarily to scale in its original version. In the ensuing *design stage*, this sketch is used as a basic for a design drawing, which already shows all the parts to scale, with connecting dimensions. At the same time general calculations are made to make sure that the design satisfies the main design criteria relating to service life, strength and cost. In the last, or *detailing stage* all the parts are finally specified, dimensioned and drawn in detail. Drawings are then made of sub-assemblies and the parts lists are compiled. For complicated designs it may be necessary to repeat one or more of these stages a number of times until the iterative process yields an optimum design.

The diagram in *fig. 2* gives a general picture of the various activities encountered in design. We see that drawing accounts for most of the time, 33%, and that 15% of the time is devoted to the actual designing. It is particularly interesting that only 3% of the total design time is used for computing. This means that very little time would be saved if the computer were only used for calculations; to obtain any real rationalization of the design process it is necessary to pay particular attention to the actual drawing and designing.

To find out where the computer could be used in the various design activities, it is necessary to separate these into *heuristic* activities and *algorithmic* activ-

Dr P. Blume is with Philips GmbH Forschungslaboratorium Hamburg, Hamburg, West Germany.

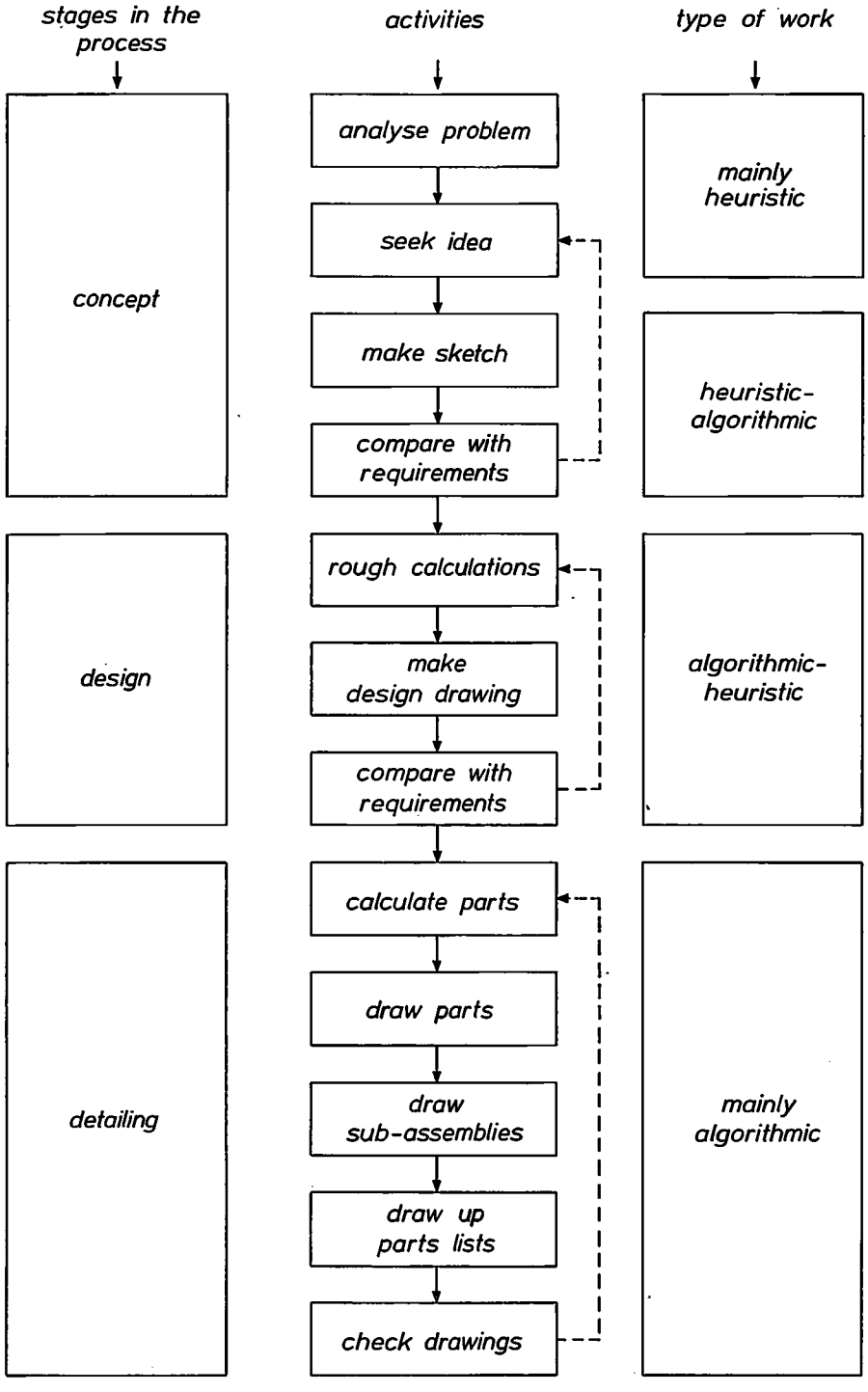


Fig. 1. Diagram of the design process for a mechanical part. The three stages of the process are shown on the left, and the detailed activities at the centre, while on the right the work is subdivided into heuristic or creative work and algorithmic work, i.e. work performed in accordance with a certain set of instructions. A rough sketch of the device is produced in the concept stage; a design drawing based on that sketch is made in the design stage. In the detailing stage the various parts are detailed and drawn.

ities [4]. Heuristic work may be defined as the creative work, i.e. 'invention', the work based on the designer's ability to think in terms of abstract ideas and to take decisions in complex situations. Except in a few elementary cases this work cannot be taken over by the computer in the present state of technology. Algorithmic work may be defined as work for which an 'algorithm' can be given, i.e. a series of instructions or procedural steps for the solution of a specific problem. Such activities can very readily be performed by a computer. In design work the actual drawing (and calculations, of course) belong to this category.

If we now apply this to the various stages of the design process as represented in fig. 1, we see that the

[1] This definition is taken from the Brockhaus encyclopedia, 17th edition, 1967.

[2] S. Hildebrand, Feinmechanische Bauelemente, Carl Hanser Verlag, Munich 1968, p. 27.

[3] U. Baatz, Bildschirmunterstütztes Konstruieren, Funktionsfindung, Prinziperarbeitung, Gestaltung und Detaillierung mit Hilfe graphischer Datenverarbeitungsanlagen, thesis, Rheinisch-Westfälische Technische Hochschule Aachen, 1971.

[4] See the thesis mentioned in note [3].

concept stage chiefly comprises heuristic work. A few algorithmic activities occur in the design stage while the detailing stage consists largely of algorithmic work. It follows that in the first instance computer-aided design can be usefully employed only in the detailing stage. To what extent parts of the design stage will lend themselves to 'computerization' depends largely on the type of design.

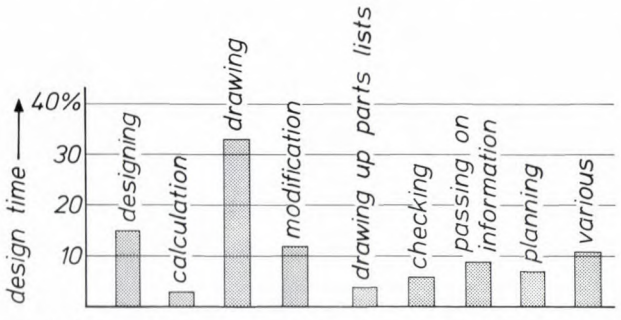


Fig. 2. Diagram showing how the total design time for mechanical parts is subdivided, according to a study made by the machine-tool laboratory of the Rheinisch-Westfälische Technische Hochschule Aachen [5]. The actual designing accounts for 15% of the time, and calculation only 3%. The largest part of the time, 33%, is devoted to drawing.

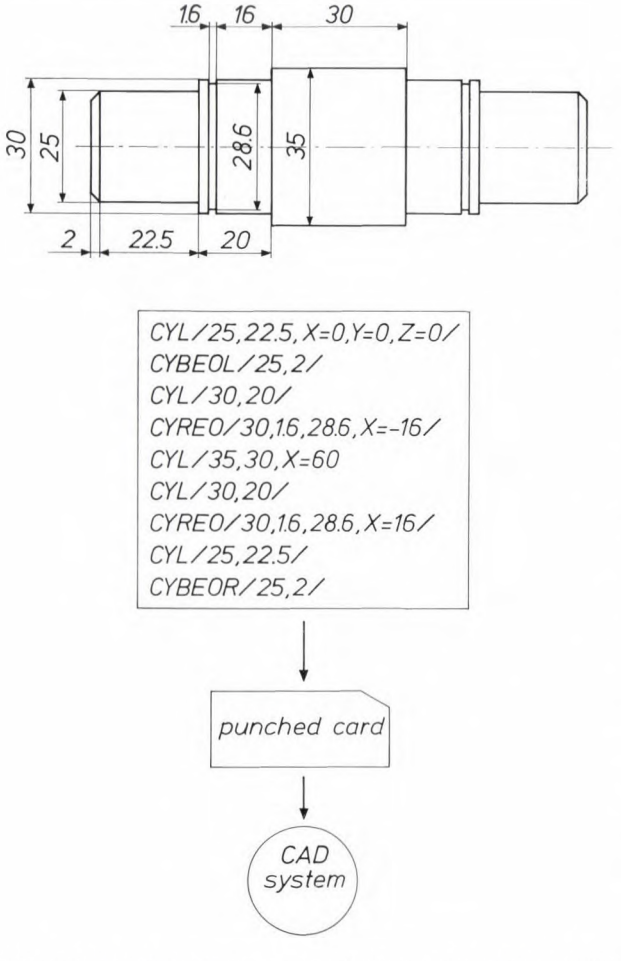


Fig. 3. Input of the shape of a workpiece to the computer. The code, taken from a workpiece-descriptive programming language [7], consists of a mnemonic description of each element of the workpiece, followed by a number of digits indicating the dimensions of the element and its locations in space. Thus CYL/25, 22.5, $X = 0$, $Y = 0$, $Z = 0$ means a cylinder with a diameter of 25 cm and a length of 22.5 cm whose axis lies along the x-axis and with the centre of its left-hand side face at the point (0,0,0). CYBEOL means 'a cylinder with a bevel on the outside, left' and CYREO 'a cylinder with a recess on the outside'.

Three main types of design may be distinguished: new designs, variants and adaptations. In the case of new designs the result is a completely new product. In such cases, therefore, recourse cannot be had to design data for earlier products; the work has to pass through all the stages of the design process. In variants the principle and operation of the design are already known so that the concept stage can be omitted. Changes may however be made to certain parts of the design. In adaptations the design of the product is already established in such detail that it may only be necessary to alter a number of dimensions.

In the following there will be a discussion of a CAD system that can be used to develop designs of the first type, i.e. new designs. (This does not of course exclude its use for the other types of design.) New designs are very often necessary in mechanical parts of electrical and electronic products; the rapid developments here mean that products and their mechanical parts are constantly subject to change. Since a new design has to pass through all the design stages, it follows from what has been said above that at present the use of a CAD system for this type of construction has to be restricted to the detailing stage. The emphasis here is on the description of the geometrical shape of the various parts of the design, to provide a starting point for making drawings and design changes with the aid of the computer.

Application of workpiece-descriptive programming languages

The geometry of the workpiece can be described with the aid of workpiece-descriptive programming languages that have been developed, at various centres, from the APT language for programming numerically controlled tools [6]. Using a shaft as our example it will be shown how the description of a workpiece is effected with a language of this kind; the language used here was developed by E. Schnelle at the Technische Universität Berlin [7]; the drawing of the shaft and the resulting text are shown in *fig. 3*.

The shaft is first divided into basic shapes or bodies such as, in this case, cylinders, which in their turn can be modified by 'form elements'. Each basic shape and each form element is described by a term consisting of a mnemonic description (up to six letters), followed by a number of digits describing the size of the body or element or its location in space. In the example in fig. 3 the left-hand part of the shaft is described first: CYL means cylinder; this designation is followed in

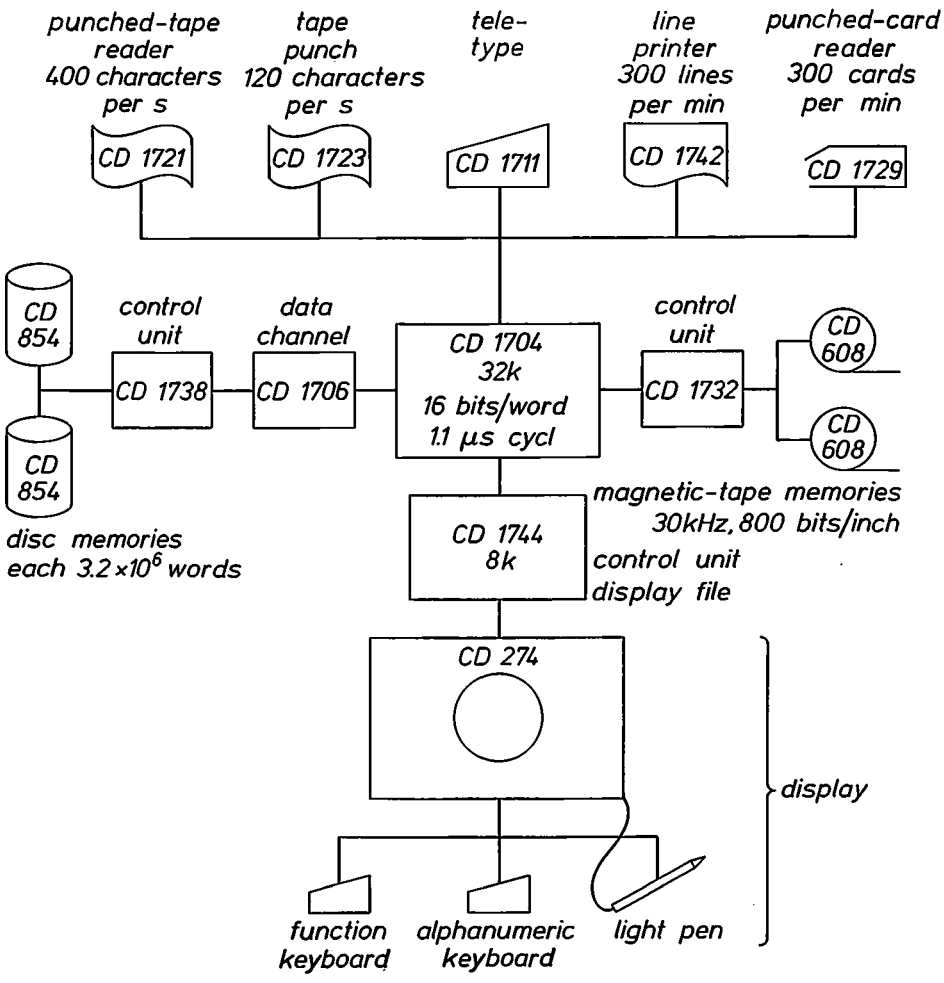


Fig. 4. The CDC 1700 Digigraphic computer system for the graphic processing of data. The interactive display is connected to the CD 1704 central processor via a control unit with a 'picture store'; the computer itself has the usual mass stores and peripheral equipment. Information from the computer store can be displayed on the screen of the picture tube and can be altered or added to by using a light pen and keyboards connected to the interactive display. This means that very fast input of both alphanumeric and graphic information is possible, while the input can be verified immediately on the screen.

turn by the diameter and length of the cylinder and the initial coordinates required to indicate the cylinder's location in space. As in APT, abbreviations are used to describe the elements; for instance, CYBEOL means 'cylinder bevel outside, left' and CYREO stands for 'cylinder recess outside'.

The description of the workpiece data by a programming language has the advantage that the user of the system can describe the workpiece at his own pace in advance without having to take computer times into account. The text is punched on to cards and only then fed to the CAD system. This method, however, has the great disadvantage that the designer, who generally expresses his ideas in a direct graphical way, has to express himself in a language with an extensive vocabulary. Moreover, this verbal system is sensitive to errors. An input system in which the designer communicates graphically is therefore preferable.

In this article the development of such a CAD/CAM system will be discussed. The system is not only used for designing workpieces, it also carries out production-planning activities right up to the generation of the perforated tapes for the manufacture of components on numerically controlled machine tools [8].

Input via an interactive display

The graphical communication with the computer can be realized in a very direct manner with the aid of an interactive display. The term 'interactive' indicates that the display permits a dialogue between the designer and the computer [9]. *Fig. 4* gives a general impression of a computer system designed for this purpose; it consists mainly of a central processor with the usual peripheral equipment (typewriter, card and tape punches and readers, a high-speed printer, and magnetic tape and disc stores) plus some other input and output devices for graphic information, such as a digitizer, a device which can transfer a drawing in digital form, e.g. on to paper tape, and a data plotter. The heart of

[5] R. Simon, Rechnerunterstütztes Konstruieren, eine Möglichkeit zur Rationalisierung im Konstruktionsbereich, thesis, Rheinisch-Westfälische Technische Hochschule Aachen, 1968.

[6] J. Vlietstra, The APT programming language for the numerical control of machine tools, Philips tech. Rev. 28, 329-335, 1967.

[7] E. Schnelle, Rechnerunterstütztes Konstruieren im Dialog, thesis, Technische Universität Berlin, 1972; see also: I. C. Braid, Designing with volumes, thesis, University of Cambridge, 1973.

[8] This research was carried out with support from the government of the Federal Republic of Germany.

[9] See the thesis mentioned in note [3].

the system is the display unit, with either a circular or a rectangular picture tube, and a picture store in which the information to be reproduced on the screen is stored in a digital code. With the aid of this store the picture is generated approximately 40 times per second so that a stationary and flicker-free picture is obtained (a 'refreshed display').

The system is made interactive by the electronic light pen, which identifies any picture element it is pointed at; see *fig. 5*. The tip of the light pen contains a photocell that picks up a light pulse from the screen at the instant when the indicated element is written by the cathode-ray beam of the picture tube. Identification consists in determining which picture element from the picture store is read out when the light pulse appears. This information is then fed to the computer, which then acts on this information, e.g. by carrying out a particular operation on the picture element, as indicated by the program. The light pen can also be used to start a program or branches of programs by pointing it at words written on the screen, e.g. instructions. To ensure that it is in fact the desired word, point or line segment that is fed in, the element pointed at always flickers initially and the information is not passed on to the computer until the user has confirmed that it is the element intended. This confirmation is also effected with the light pen, which is pointed at a special word, 'accept', on the picture tube.

The method of inputting a geometrical shape 'interactively' via the display will now be described, again using the example of fig. 3. The drawing of the workpiece will appear within a rectangle. On a circular picture tube four segments remain outside this area and these are available for writing in lists of instructions and comments. These lists represent choices of actions that the user can select with the light pen, and are called *menus*.

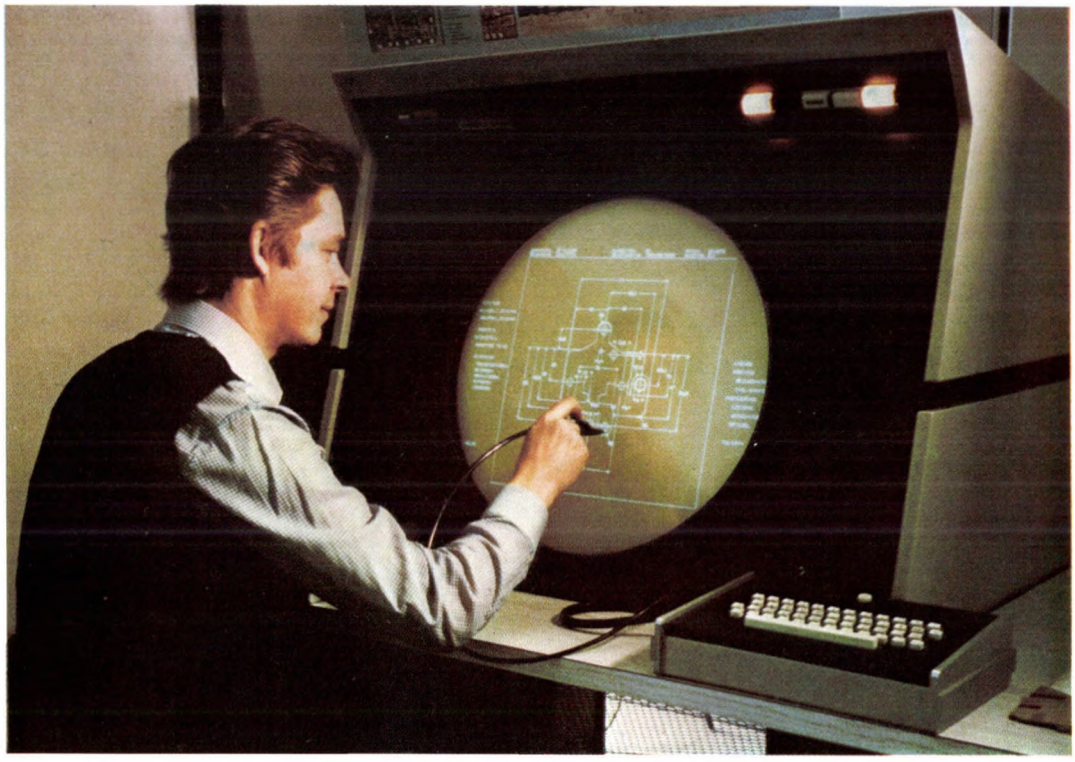


Fig. 5. The interactive display in operation. If the light pen is pointed at a particular element on the display, the position of that element is conveyed to the computer. The display can be used to input graphic information to a computer. It is also possible to intervene in the progress of a program by pointing at command words displayed on the picture tube. The user can thus engage in a dialogue with the computer via the display; hence the term 'interactive'.

The dialogue between the computer and designer starts with the presentation in one of the four segments of the picture tube of a menu of workpiece families: turned parts, sheet-metal parts, milled parts, etc. Selection of one of these words with the light pen indicates to the computer that the program for describing, say, turned parts has to be run. A menu of elements describing such parts then immediately appears on the screen. Since there is not much room in the segments and research has shown that recognition of a word becomes disproportionately more difficult with increasing length

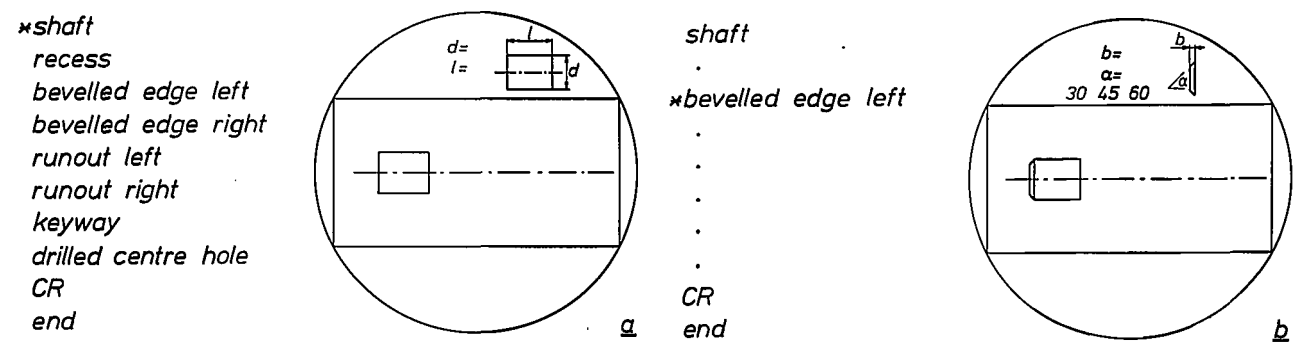


Fig. 6. Dialogue between the designer and the computer during the interactive input of the geometry of a shaft. The central rectangular region of the circular screen contains a drawing of the part of the workpiece that has already been input. Lists of the names of elements descriptive of the kind of workpiece desired can be displayed in the four segments round this rectangle. These lists are known as 'menus'. When one of these elements is indicated by pointing the light pen at it, this element is added to the drawing of the workpiece. The dimensions are inserted via a keyboard or by pointing at a 'number' menu that is also displayed on the screen. *a*) When a shaft is to be described the word 'shaft' is assigned from the menu on the left and the diameter and length are input. This part of the shaft then appears in the working field of the picture tube. The following are then added successively in the same way; a bevelled edge (*b*), a recess for a retaining ring (*c*) and a machine-tool runout (*d*). The shaft is now extended towards the right by the addition of a second shaft section (*e*) to which a keyway is added (*f*).

of the list, a menu usually contains no more than eight words. If more than eight words are necessary, they are divided between a number of hierarchically arranged menus.

The first menu for turned parts contains the most important shapes and form elements that can occur, such as a shaft section, a bevel, a recess, a tool runout, etc. In describing a turned part of the type shown in fig. 3 the workpiece is divided into separate shaft sections of a certain thickness and then, after the starting point has been determined, the first section is input. The designer therefore first points to the word 'shaft' on the menu (*fig. 6a*). Each time a particular element has been selected the parameters required for its further definition appear in the top segment of the picture tube, together with a drawing to clarify them. In the case of the shaft this drawing is a rectangle and the parameters are the diameter and the length. The values of the parameters are input either via a keyboard or with the light pen by pointing at digits in a menu displayed on the screen and consisting of the digits 0 to 9. The advantage of this method is that the designer does not have to put the light pen down to use the keyboard. When the dimensions have been input, a drawing of the desired section of the shaft appears at the indicated position in the rectangle, symmetrically in relation to the axis (see fig. 6*a*).

The end of this shaft section must now be bevelled. The term 'bevel' is then pointed at and another diagram appears, this time with the angle and width of the bevel

as parameters, together with a small menu of the most common values for the angle (fig. 6*b*). Displaying these angle values has the advantage that it is only necessary to point at the desired value instead of supplying each digit separately. From this rather trivial example we can see how the system ensures that the decisions that the user has to take are prepared in such a way that they demand the least possible effort from him. This emerges still more clearly in the steps which follow: the introduction of a recess for a retaining ring and a runout for grinding. After indicating the term 'recess' (fig. 6*c*), another diagram appears on the screen complete with parameters; the dimension l of the new element where it connects with the part already formed is particularly important. Since there are standard thicknesses and widths for these retaining rings and recesses, depending on the diameter of the shaft, these dimensions are read directly from tables stored in the computer memory. The designer therefore only has to indicate the parameter l.

Four shapes A, B, C and D are laid down in the standard DIN 509 for the tool runout (fig. 6*d*). By pointing at these letters one by one the designer can now make diagrams of these four possibilities visible on the tube, enabling him to select the shape which he finds most suitable. Since the parameters of the four shapes are specified for each shaft diameter in the standard, the designer only has to indicate the shape he has selected and confirm it.

The following stages in the input of the shaft take place in an analogous way (figs. 6*e* and *f*). We shall only consider one more possible stage in more detail: the addition of a keyway. Like the tool runout, this element is covered by a DIN standard; in any particular company a preselection is generally made from the standard versions and these are listed in a menu. The length of the keyway can still be chosen freely within certain limits; this dimension depends on the torque to be transmitted. Thus if a program is included that calculates the length of the key from the maximum torque required, all the user of the system has to indicate in addition to the torque is the connecting dimension (l in fig. 6*f*) in relation to one of the end faces of the shaft.

In the example given above we see that, despite the use of the computer and the interactive display, a considerable amount of data still has to be input manually, even for a simple workpiece. The example also shows clearly that by the start of the detailing process, i.e. at the point where the CAD system begins, the workpiece has already assumed a fairly definite shape. At this point it might not really seem that the use of this equipment is worth while in comparison with conventional drawing on a drawing-board. If the drawing alone is considered, then it is definitely not very helpful. But the great advantage of using the computer, however, is that when the dialogue with the computer has been completed a full geometrical description of the workpiece is available in the computer memory. This allows the information to be automatically processed, in whole or in part, from this point onwards for the production-planning or preparation processes that follow the detailing. These activities can also be executed with the interactive display: the workpiece data can be stored in an electronic file (on magnetic tape or magnetic disc); immediate access is then available if updating is necessary; the parts of the workpiece can be called separately from the file and arranged on the screen to form subassemblies. Alternatively, the data can be used again in a dialogue, for generating the control information for producing the parts on numerically controlled machine tools.

The geometrical structure of a workpiece

We shall now take a closer look at the geometrical model of a workpiece. The shape of the body can be described unambiguously by giving the positions of all the points on its bounding surfaces. If these surfaces have a completely irregular shape, it can mean in the extreme case that an infinitely large number of points and hence an infinitely large amount of data is necessary for an exact description. Usually, the designer will look for a more regular shape with more regular limiting surfaces, which gives a good approximation to the body within the specification and can be described with a smaller amount of data.

Engineering workpieces are not usually irregularly shaped bodies but objects made on machine tools and hence of a more regular shape. Turned parts, for instance, are made on a lathe and have rotational symmetry. An investigation carried out in the mechanical engineering laboratory of the Rheinisch-Westfälische Technische Hochschule Aachen, on workpieces from a production run [10] shows that a large number of the workpieces examined can be made up from a small number of basic shapes such as blocks and cylinders. By way of example, *fig. 7* shows a shaft and the basic shapes into which it can be divided. The workpiece-descriptive programming languages are based on this principle. Since, as we shall show below, basic shapes can be described with only limited data, the description of objects composed from them, modified if necessary by simple form elements, requires a minimum of information.

A basic geometrical shape is determined by a hierarchy of descriptive elements; *fig. 8* shows this hier-

archy for a cube. On the top step we have the body itself; on the second step are the surfaces bounding the body; these in turn are bounded by the edges that form the third step and these are in turn bounded by the

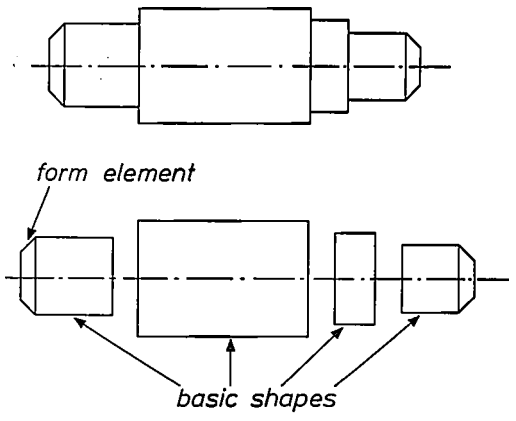


Fig. 7. Example of a composite body and the basic shapes, in this case cylinders, into which it can be divided. The basic shapes, in turn, can be further modified by the addition of form elements such as bevels, as shown here.

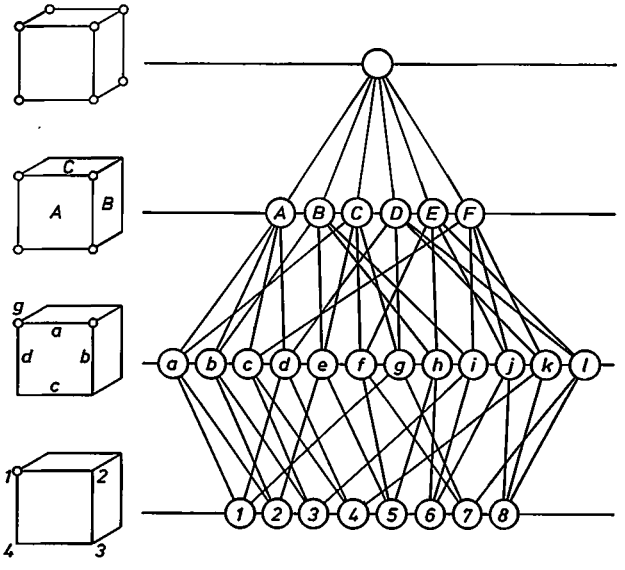


Fig. 8. Schematic representation of the geometrical structure of a basic shape, in this case a cube. The structure can be described with a number of hierarchically arranged lists in which the various elements such as faces, edges and vertices are given, and with descriptions of the relationships between these elements. The division into elements and relationships is very convenient for the storage of the structure in a computer memory.

points forming the fourth step. From this classification there emerges a hierarchy of lists for describing an object, e.g. an 'edge list' comprising edges *a* to *l* and containing the description of these edges (straight, circular, etc.) and a 'point list' containing the coordinates of the various points. In addition to the elements themselves, the relations between the elements can be described. In the case of the cube, for instance, there is a relation between the surface *A* and the edges *a*, *b*, *c* and *d*, since *A* is bounded by these edges; there is also, for example, a relation between vertex *1* and edges *a*, *d* and *g*, because vertex *1* is the vertex common to these edges.

For basic shapes a complete description of the geometry using these lists is in general highly redundant. A cube, for example, can be described adequately both in size and its position in space by giving the length of its side together with the direction cosines of one of the surfaces and the coordinates of a single vertex point. There is, however, considerably less redundancy in the more complicated composite objects generally encountered in practice. However, practice shows that considerable loss of time in the dialogue via the interactive display is incurred if all the coordinates of the angular points and centres of the basic shapes and all the data on the edges have to be generated before the workpiece can be displayed. It is therefore advisable to store a complete 'redundant' geometrical description of the workpiece together with its basic shapes and form elements in the computer memory.

Data storage in the computer

The geometric data is stored in a computer system in the form of 'data structures', i.e. structured collections of data in which various objects and their interrelations are recorded. In recent years a number of data-structure types have been developed; one of these is that of the CODASYL group [11]. This design was developed on the basis of experience with the storage of data for commercial data processing. In our experience, however, the CODASYL structure is also very useful for the storage of geometrical structures.

The elements of the CODASYL structure are the *record* and the *set*. A record is a representation of an object by its properties or attributes. Thus the record of the object 'point' consists of the serial number of the point and its *x*-, *y*- and *z*-coordinates. A group of objects with the same characteristics constitute a record type; the group of all point records thus forms the 'points' record type. Conversely, each record individually is called an 'occurrence' of a record type, which means to say that it is one of the class forming the record type. A set defines a relation between record types, e.g. between the record type of a body and the record type of its bounding surfaces. The term 'occurrence' is used in connection with a set, in the same way as in connection with a record type, if a particular

[10] See the thesis mentioned in note [4].

[11] CODASYL Data Base Task Group, report, April 1971; CODASYL is an acronym for COnference of DAta SYstems Languages.

relation between individual records is intended. In practice the set is realized by means of a reference contained in the record, for example to certain other records or to a table that in its turn contains references. The following notation has been developed by C. W. Bachman [12] for representing data structures: record types and records are shown as rectangles and a set is represented by an arrow connecting the rectangles (*fig. 9a* and *b*).

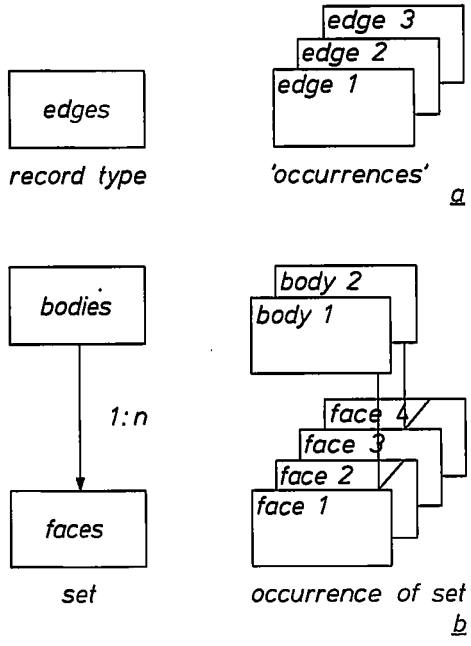


Fig. 9. Schematic representation of the basic elements in the CODASYL structure for the storage of data in a data bank. *a*) A rectangle represents a *record*, a description of an object, or a *record type*, i.e. the collection of all the records of a particular type. Every record can be regarded as an occurrence of a record type. In the application described here the record is used to describe an element in the geometrical description of a workpiece. *b*) An arrow between two rectangles represents a *set*, the description of a relationship between two record types. The set is defined such that there is always one record that is the OWNER of one or more MEMBERS (e.g. a single body consisting of two (curved) surfaces); only $1:n$ relationships can be described by sets. The term *occurrence* is also employed in the case of a set if what is meant is the relation between an OWNER record and its MEMBER records.

In the CODASYL structure the set is organized in such a way that a single record type is always shown as the OWNER, one or more record types as MEMBERS. The 'occurrence' of a set always consists of one OWNER record and zero, one or more MEMBER records (e.g. a body as the OWNER and two or more bounding surfaces as MEMBERS). This means that a set can only define $1:n$ relationships. The arrow representing the set always points from the OWNER to the MEMBER. In the geometrical data for which we wish to use the CODASYL structure, $m:n$ relationships occur regularly instead of $1:n$ relation-ships, e.g. between edges and vertices: each edge has two end points and each vertex can belong to two or more edges. If, in an 'edge-vertex' set, we started from a vertex as MEMBER, two edges would again occur as OWNERS of this vertex, which is not allowed in the CODASYL structure. This problem can be overcome by introducing *relation records*, which are extra records used only as links between other records. By way of example *fig. 10a* and *b* show how the connection between edges and vertices can be defined with the aid of relation records. In the 'edge-relation record' set each OWNER record 'edge' has two 'occurrences' of the relation-record type as MEMBERS. Each of these records in turn is associated with another relation record as MEMBERS of a vertex as OWNER in the 'vertex-relation record' set. To find the vertices associated with a particular edge we must therefore look first in the 'edge-relation record' set of MEMBER relation records associated with the edge as OWNER; the vertices associated with these relation records can then be found as OWNERS in the 'vertex-relation record' set.

With the aid of these elements of the CODASYL system the data structure of engineering workpieces can be represented in the manner illustrated in *fig. 11*.

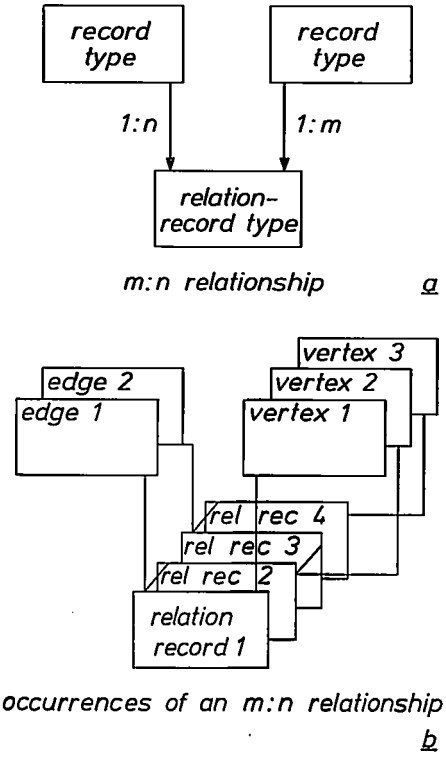


Fig. 10. *a*) Representation of a data structure in which an $m:n$ relation occurs. To enable the records and sets as defined in fig. 9 to be employed in this case, a relation record forming the link between the two records is inserted. *b*) The examples given here are the occurrences of two sets that indicate, via a relation record, the connection between the edges of a cube and three vertices, one of which is the common vertex.

To distinguish them from ordinary records the record relation types in this figure are symbolized by circles.

While data structures such as the CODASYL structure are adapted to the requirements of data processing, the form in which they are stored in a computer

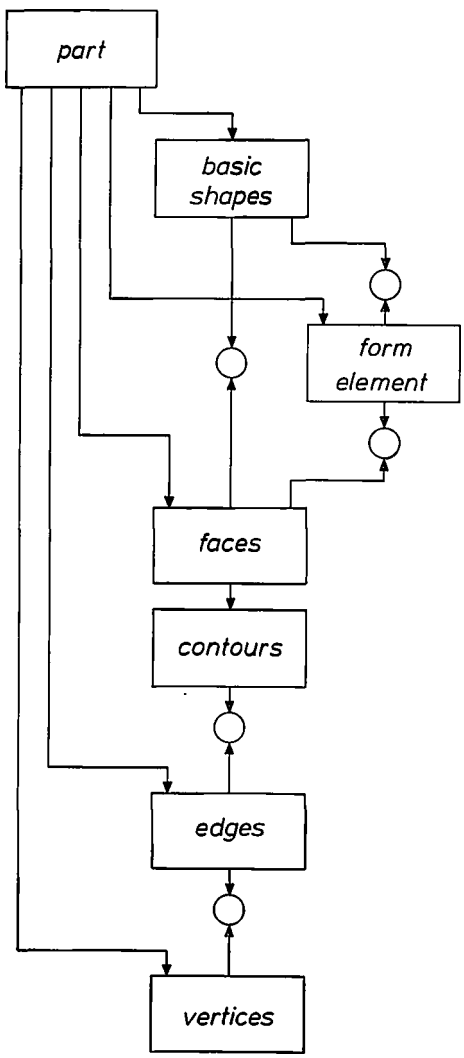


Fig. 11. Representation of the data structure of a workpiece in the CODASYL system. Relation records are shown here as circles.

memory has not yet been dealt with. This form, the *storage structure*, is not only influenced by the data structure but also by the organization of the memory, e.g. the addressing, the coding of information, etc. This point cannot however be discussed in this article. Knowledge of the storage structure is not in fact required by the user; the data structures expressed by means of CODASYL data-structure elements described above can be stored in any data bank based on the CODASYL system (e.g. the Philips PHOLAS [13]) without further knowledge of the storage structure. A data bank of this kind is used to store the geometrical data for mechanical workpieces.

The integrated CAD/CAM system

For executing the various activities required in the designing and production planning of sheet-metal parts, we have developed an integrated CAD/CAM system, which is shown schematically in *fig. 12* [14]. The application of the system starts in the detailing stage of product design and covers tool design and the programming of numerically controlled wire spark-machining equipment used for making the production tools.

The geometrical data for sheet-metal parts is input in the same interactive way as described above for turned parts. Our experience is that much less use can be made of form elements here. Also, fewer design rules are available. Since sheet-metal parts are much used in assemblies such as chassis, cabinets and frames, it is likely that design systems for these will be developed in the future.

Once the geometry of the workpiece has been described, a digital representation is stored in the parts file. The technological and organizational data are then added and the part is dimensioned (*fig. 13*). The dimensioned part must of course be clearly readable, for example the dimension lines and the lines of the part must not intersect. Producing such a drawing is usually so complicated (see *figs. 14* and *15*) that this activity cannot be performed by an algorithm but requires human intervention: the light pen is pointed at the line that has to be dimensioned. This line then starts to flicker as was explained above, and when it has been confirmed that it is the line intended, the coordinates of its end points are read from the data structure, the dimension is calculated and then automatically displayed on the screen together with the dimension and extension lines. If some of the dimensions thus generated are inconveniently located, e.g. because they intersect other dimension lines or contour lines of the workpiece, they can be moved to any other more suitable place by the light pen. To enable the representation of the dimensions to be called later, together with the geometry of the workpiece, this information is also stored in the parts file.

Drawings of the workpiece design can be made automatically with an x-y plotter; this data is supplied by the parts file via a postprocessor that converts the data from the data structure into machine code for the plotter pen. It is possible to define a standard interface so

[12] C. W. Bachman, Data structure diagrams, Data Base 1, No. 2, 1969.

[13] Introduction to PHOLAS, publication of Philips Electrologica BV, Apeldoorn, July 1974.

[14] We should like to acknowledge the valuable help, relating to the mechanical engineering of the system, that was given by Dr Burmester of the Philips Data Systems Division, Eiserfeld, West Germany.

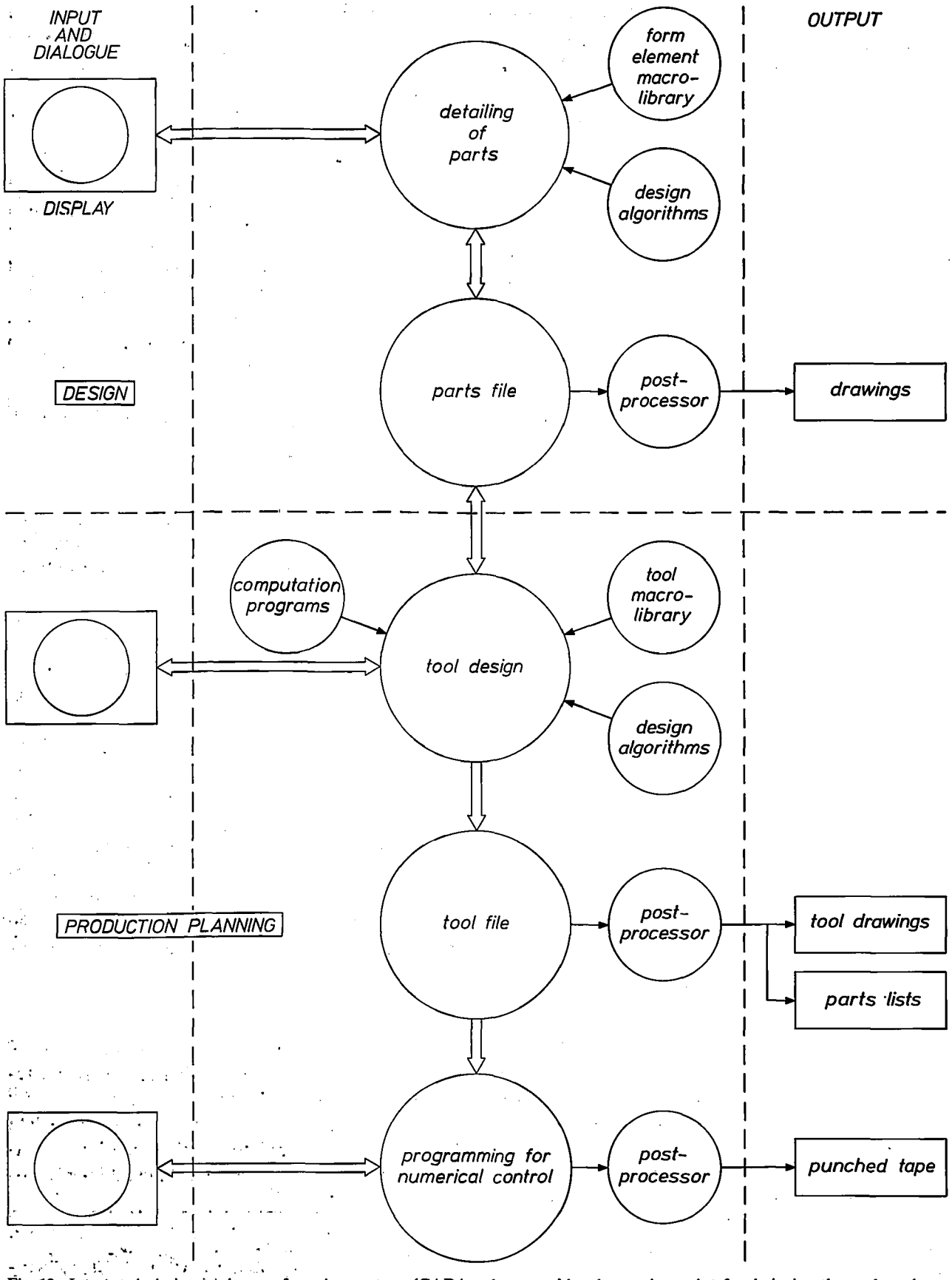


Fig. 12. Integrated design and manufacturing system (CAD/CAM) for punched metal parts. The upper part shows how the detailing stage of designing takes place in a dialogue between the designer and the computer via the interactive display. The workpiece data generated in this way is stored in the parts file. This data provides the starting point for designing the tools such as single-stage blanking tools, again by means of the interactive display. The resulting data is collected in the tool data bank. Finally, in the control-programming stage, information is added in a dialogue for generating the numerical-control tapes.

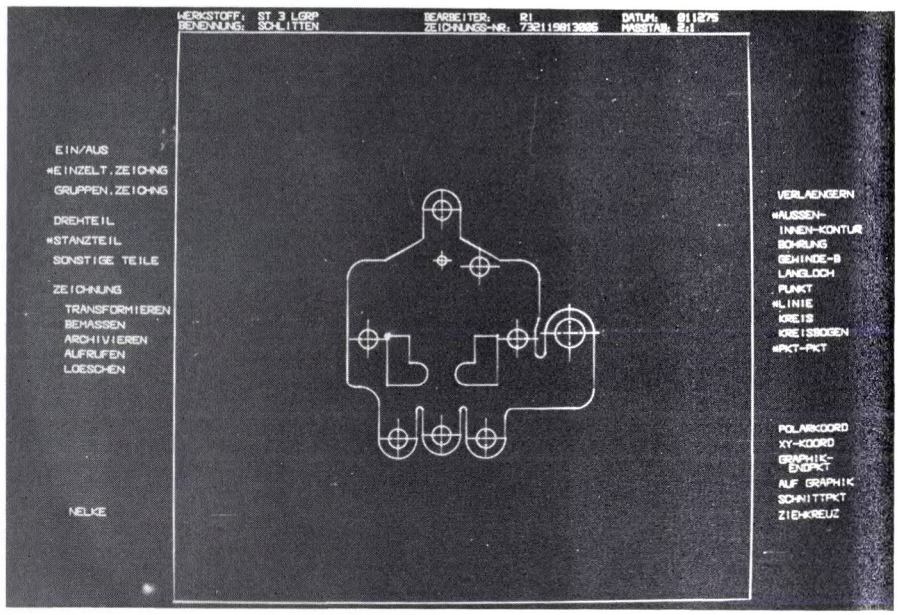


Fig. 13. Drawing of a punched metal part displayed on the screen after input of the geometry. The description is now complete but no dimensions have so far been added.

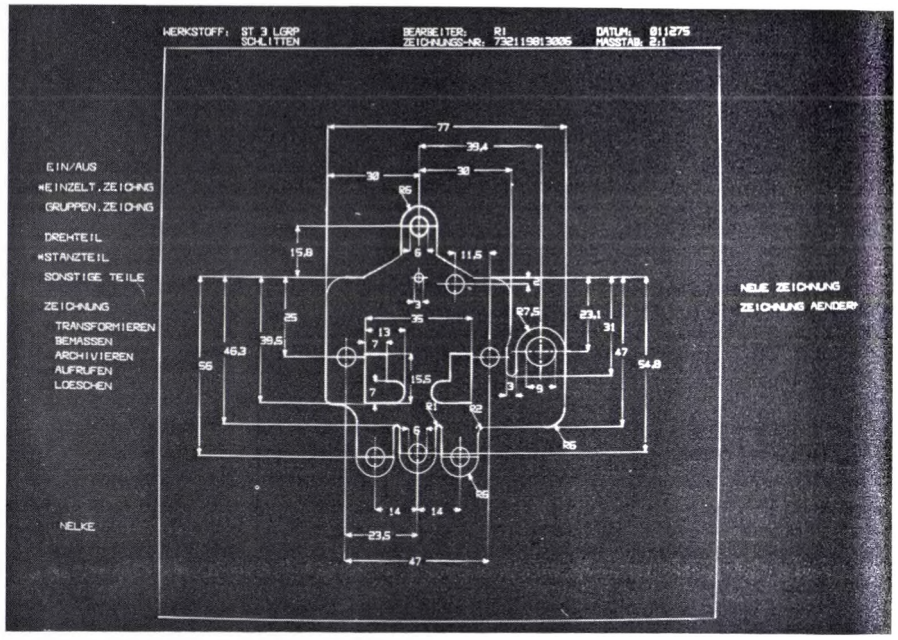


Fig. 14. The same workpiece as in fig. 13 after dimensioning.

that the various makes of x-y plotter will operate with only one program for converting the data structure to that code.

In the integrated system shown in fig. 12 the parts file forms the interface between the activities of the design and production-planning offices. In the design of sheet-metal parts a typical activity in production-planning is the design of the punching tools required; this is also done with the aid of the interactive display. The part is called from the data bank, and then the outer contour and the inner contours are identified with the light pen. These contours are nearly identical with the cutting contours of the main tool parts — the die and punch. A computer program now calculates

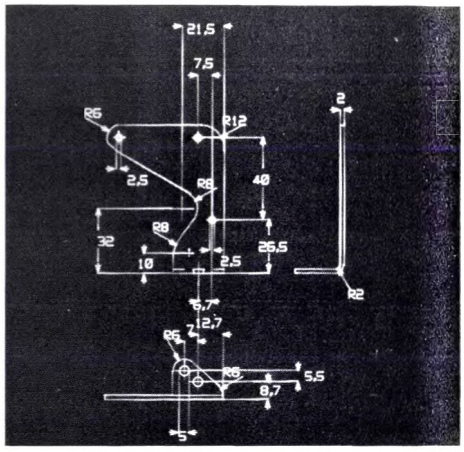


Fig. 15. Drawing of a bent sheet-metal part as displayed on the screen; three views are given, together with the dimensions.

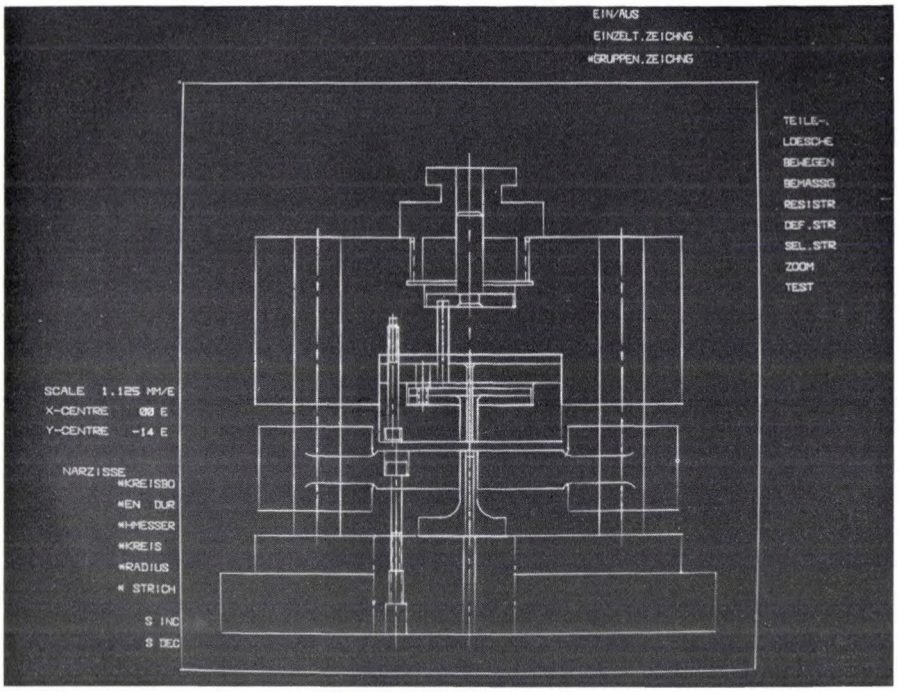


Fig. 16. Drawing of a punching tool as described in a dialogue with the computer.

the total length of these contours and then determines the cutting force, taking into account the shear strength of the material. For parts that require bending, a 'development' on a plane must first be made. This is also done interactively, by showing the various views of the part on the screen to form the developed part. The corrections that have to be applied in determining the dimensions of the part on account of the distortions occurring along the bent edges can be carried out by the computer, e.g. in accordance with the recommendations of the standard DIN 6935.

Once the cutting force has been determined, the size of the tools is determined. Then the contour is positioned in the die; where possible the point of application of the cutting force is made to lie at the centre of the tool; this positioning operation is carried out automatically by a computer program. After the angle at which the strip of material is fed to the press has been selected, the contour can still be rotated round this point of application to a position in which the least possible loss of material occurs in the punching operation.

When this has been done, the shape of the punch and the various piercing tools (for the smaller holes) can also be determined by using design algorithms. In addition, other tool parts are detailed, such as the 'stripper' and the 'ejectors', which, after the punching, remove the material. The parts are then dimensioned interactively and the data stored in the tool data bank (see fig. 12). Drawings are made in the same way as for the parts. *Fig. 16* shows an example of a punch tool as displayed on the screen.

The design of the tools is followed by the programming for producing these tools on numerically controlled machines. We assume that a numerically controlled wire-type spark-machining equipment — in which a wire electrode is moved through the material in much the same way as a fretsaw — is available for cutting the contours. A machine of this kind can be programmed very efficiently in the same interactive manner used for input of the geometrical data. The tool is called from the tool file and displayed on the screen. The contours to be cut are then identified with the light pen and the machining movements and the non-machining movements (the displacement of the wire to a new starting position) are indicated in relation to the initial hole for the insertion of the spark wire; see *fig. 17*. Since sharp corners cannot be cut with the wire spark machine, any corners have to be rounded off. The numerical-control tape is then generated automatically by means of a postprocessor. Information about the tool path can also be stored in a data file for any later processing necessary (this is not shown in the figure).

The programs relating to part description were completed in October 1975, and programs for designing single-stage blanking tools are approaching completion. A postprocessor for programming the numerical control has been made for the 'AGIE-Cut' spark machine type DEM 15, and the dialogue for this machine is also ready. A number of tools have already been machined with the aid of a tape prepared in this way. The first prototype of the complete system will be ready by the end of 1976.

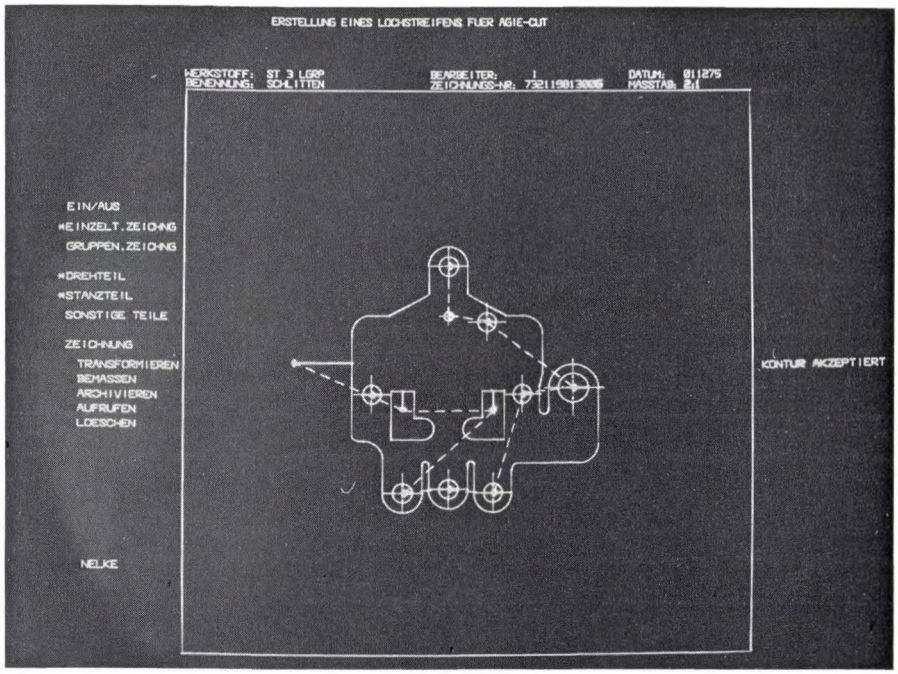


Fig. 17. The programming of the numerically controlled wire-spark machining equipment, for cutting a punch pin for the part shown in fig. 13. The shape of the punch is called from the tool data bank; the designer then uses the light pen to indicate the path the wire must follow through the material, starting from a new hole each time. The movements between the various starting holes are indicated by dashed lines. The path to be followed by the wire is recorded on punched tape.

The economic aspects

In conclusion, we should like to consider the economic aspects of the use of an interactive display. The system discussed here is based on the CDC 1700 Digigraphic computer system; the heart of the system is a minicomputer with a store for 32 k words, each of 16 bits, and a cycle time of 1.1 μs (fig. 4). At present the hourly rate for a system of this kind used 12 hours per day is still about twice the cost of one man-hour. If, however, the continuing development towards cheaper hardware and increasing personnel costs is taken into account, the time can be seen to be rapidly approaching when an hour of computer time will no longer be as expensive as a man-hour. When that time comes, design work with the CAD/CAM system will only have to be completed twice as quickly as in the conventional way for economic justification. If the full potential of the system is considered, it will be found that this factor can already be attained. Economically, therefore, the system will probably not give rise to any problems in the next few years but the number of programs available will have to be considerably enlarged.

Summary. Description of an integrated computer system for designing mechanical parts. The vital feature of the system is the 'interactive display' that enables the user to feed graphical information into the computer by means of a light pen. The application of the system is as yet restricted to the detailing stage of the design process; the method consists of a dialogue between the designer and the computer via the display, in the course of which the designer can use the light pen to select from a number of basic shapes and form elements presented on the picture tube, thus building up the workpiece step by step. All dimensions are fed in by keyboard or with the light pen. The article begins with a short survey of the conventional design process and then, as example to illustrate the use of the CAD system, describes the input of the data for a turned part via the interactive display. The descriptive method employed for this, consisting in the use of basic shapes and form elements, is taken from the workpiece-descriptive programming languages. The CODASYL structure developed for commercial data files is used for description of the workpiece and for storage of the workpiece data in the computer memory. In this process the geometrical structure of the workpiece is described by a hierarchy of the geometrical elements (faces, edges, vertices) and their interrelationships. These can be represented by the 'record' and 'set' structure elements of the CODASYL system. In the integrated CAD/CAM system for punched metal parts, which is being developed at the Philips Hamburg laboratories, the punching tools required are also designed. First the cutting force required is computed and the most economical strip layout is determined, again via the display. The die plate and punch are then designed. This is followed by the numerical-control programming for the manufacture of these tools. This process, also carried out on the screen, yields the punched tapes used in this case for the numerical control of a spark-machining set. The article ends with a brief survey of the economic aspects of the system.

Recent 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*
Mullard Research Laboratories, Redhill, Surrey, England	*M*
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*
MBLE Laboratoire de Recherches, 2 avenue Van Becelaere, 1170 Brussels (Boitsfort), Belgium	*B*
Philips Laboratories, 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A. (by contract with the North American Philips Corp.)	*N*

Reprints of most of these publications will be available in the near future. Requests for reprints should be addressed to the respective laboratories (see the code letter) or to Philips Research Laboratories, Eindhoven, The Netherlands.

E. Arnold & M. Poleshuk: Carrier generation at the Si-SiO_2 interface under pulsed conditions.
J. appl. Phys. **46**, 3016-3018, 1975 (No. 7). *M, N*

F. Bagdasarjanz: Quantisiert adaptive Entzerrung von Fernsprechleitungen zur Datenübertragung.
Thesis, Zürich 1975. (Philips Res. Repts. Suppl. 1975, No. 9.) *E*

D. Bois & D. Beaudet: Photoluminescence study of the shallow acceptor states in *n*-type GaAs.
J. appl. Phys. **46**, 3882-3884, 1975 (No. 9). *L*

J. Bootsma: Liquid-lubricated spiral-groove bearings.
Thesis, Delft 1975. (Philips Res. Repts. Suppl. 1975, No. 7.) *E*

K. H. J. Buschow, M. Brouha & C. Langereis: Spontaneous magnetostriction in $ThCo_5$.
Solid State Comm. **16**, 789-790, 1975 (No. 6). *E*

K. Carl: Ferroelectric properties and fatiguing effects of modified $PbTiO_3$ ceramics.
Ferroelectrics **9**, 23-32, 1975 (No. 1/2). *A*

D. den Engelsen & B. de Koning: Ellipsometry of black lipid membranes of egg lecithin and chloroplast extracts.
Photochem. Photobiol. **21**, 77-80, 1975 (No. 2). *E*

I. Flinn: Piezoelectric ceramics and their applications.
Physics Education **10**, 274-280, 1975 (No. 4). *M*

G. Groh: Holographische Methoden in der medizinischen Diagnostik.
Radiologe **15**, 236-244, 1975 (No. 6). *H*

W. K. Hofker (Philips Research Labs., Amsterdam Division): Implantation of boron in silicon.
Thesis, Amsterdam 1975. (Philips Res. Repts. Suppl. 1975, No. 8.)

J. G. Kloosterboer: Interaction of bivalent metal ions with their chelates of ethylenedinitrilotetraacetic acid (EDTA) and 1,2-*trans*-cyclohexylenedinitrilotetraacetic acid (CDTA).
Inorg. Chem. **14**, 536-540, 1975 (No. 3). *E*

R. Koppe: Das Abbildungsproblem beim rechnerunterstützten Entwurf von Layouts integrierter Schaltungen.
Angew. Informatik 1975, 223-232 (No. 6). *H*

J. G. M. de Lau: Influence of chemical composition and microstructure on high-frequency properties of Ni-Zn-Co ferrites.
Thesis, Eindhoven 1975. (Philips Res. Repts. Suppl. 1975, No. 6.) *E*

A. Mircea & A. Mitonneau: A study of electron traps in vapour-phase epitaxial GaAs.
Appl. Phys. **8**, 15-21, 1975 (No. 1). *L*

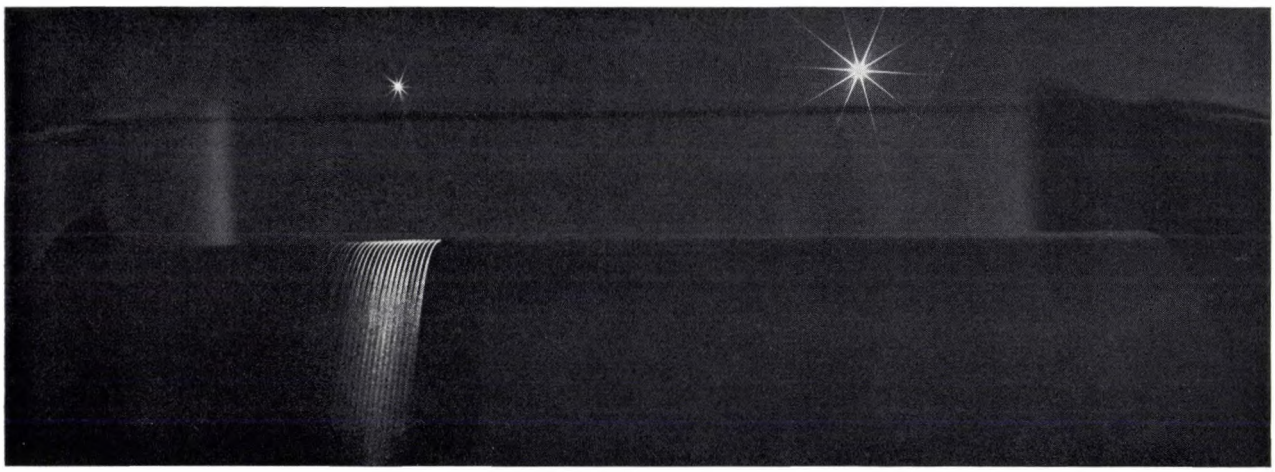
R. F. Mitchell & E. Read: Suppression of bulk wave radiation from surface acoustic wave devices.
IEEE Trans. **SU-22**, 264-270, 1975 (No. 4). *M*

A. Pirotte & P. Wodon: A query language for a relational data base.
Lecture Notes in Computer Science **26**, 524-531, 1975 (Springer, Berlin). *B*

L. Verhoeven: CAL; a videophone colour system for low bandwidth channels.
Nachrichtentechn. Z. **28**, 99-101, 1975 (No. 3). *E*

PHILIPS TECHNICAL REVIEW

VOLUME 36, 1976, No. 7



Optical communication by means of glass fibres

A new technique is gradually beginning to emerge for the transmission of information by cable. This technique can offer a very wide bandwidth, yet the quantity of material required per kilometre of cable is relatively small. The element carrying the information flow is not a metal wire or a coaxial tube, but an extremely thin glass fibre of special composition, inside which modulated light waves can propagate. The light is produced and detected by solid-state elements, whose dimensions are well suited to those of the fibres.

In this issue of Philips Technical Review examples are presented of results that have been achieved by Philips Research in a number of sectors in this field. These relate to methods for making glass fibres and assessing their suitability, the development of a stable $GaAs/Al_xGa_{1-x}As$ laser, investigating the possibilities of detection with an avalanche photodiode, a modulation method in which the intensity of the laser pulse is independent of temperature, and a solution to the difficult problem of mechanically coupling the very thin fibres to the solid-state elements. The issue starts with a short introductory article that briefly outlines the potential and the difficulties of optical communication with glass-fibre cables.

The title photograph shows part of a drum with two glass fibres wound on it; the one to the right is suitable, the other is not. The left-hand fibre loses a great deal of light sideways, whereas in the right-hand fibre (200 m long) almost all the light emerges at the end.

Optical communication systems with glass-fibre cables

K. Mouthaan

With the increasing density of telecommunication traffic transmission facilities of ever greater bandwidth are required. This necessity for greater available bandwidth arises not only from the increasing use of the telephone, but also from the desire to offer more services to the individual user. Data transmission, video telephony [1] and facsimile are typical examples of such services.

A familiar wideband transmission device is the coaxial line, in which the information is conveyed in the form of electrical signals. A single coaxial line can handle 10 000 or more telephone channels, e.g. by the well known technique in which signals for these channels first modulate carrier waves ('carriers') and are then 'stacked' in frequency. A number of lines can be collected together to form cables that are capable of carrying up to 100 000 telephone channels in both directions.

At present attention is being concentrated on an optical communication system, in which the information is transported in the form of light signals. In this system glass fibres are employed as wideband optical waveguides. Now that glasses with very low optical losses [2] are being developed, and semiconductor lasers [3] that will operate continuously at room temperature and very sensitive avalanche photodiodes [4] are becoming available, such systems are becoming of increasing practical interest. A communication system built from these elements has the essential features shown in *fig. 1*.

The glass fibres, about 100 μm in diameter, can be massed together to form thin and flexible cables that are still relatively easy to install even when the number of fibres is large. The cables are completely insensitive to electrical interference, and there is no cross-talk between the various fibres in the same cable. Compared with conventional cable systems, glass-fibre cables should be less expensive and easier to lay. In addition, copper, an increasingly scarce commodity, will be replaced by glass.

In this article we shall show how the potential applications of the optical communication system are related to the characteristics of its various components; the article is also intended as a general introduction to the other contributions in this issue of Philips Technical Review.

Dr Ir K. Mouthaan is with Philips Research Laboratories, Eindhoven.

The components of an optical communication system

In the simplest case the glass fibre consists of a core, about 50 μm in diameter, contained in a cladding of lower refractive index than the core; this brings the overall diameter of the fibre to the figure of 100 μm mentioned earlier. Light conduction in the core results from total internal reflection at the interface between the core and the cladding (*fig. 2*). For total internal reflection the angle of incidence at the interface must be smaller than the critical angle, which is determined by the refractive indices of the core and cladding. If the refractive indices differ by 1%, a value that is considered for practical applications, the critical angle is about 8°. At the end face of the fibre this corresponds to a maximum angle of 12° between the incident light rays and the axis of the fibre, because of the refraction at the glass/air interface.

The optical losses in the glass fibre directly determine the spacing at which repeaters have to be included in the cable. In only a few years it has been possible to reduce these losses to less than 10 dB/km. On a laboratory scale it has even been possible to make glass fibres whose losses are equivalent to 1-2 dB/km.



Fig. 1. Schematic representation of an optical telecommunication channel with a glass fibre *F*, a GaAs laser diode *L* and an avalanche photodiode *P*. The laser emits a beam of infrared radiation whose axis is in line with the junction of the laser diode. The avalanche photodiode detects radiation incident at right angles to the *P-N* junction.

The light source considered most suitable for this application is a GaAs laser diode, while in some cases use can also be made of an incoherently radiating GaAs diode (light-emitting diode or LED). Both sources radiate in the near infrared (0.8-0.9 μm) and can easily be modulated by varying the diode current. The diode material can be doped appropriately to ensure that the wavelength of the emitted radiation is located at or near a minimum in the attenuation of the glass [5]. A suitable detector for the near infrared range is the very sensitive Si avalanche photodiode. The permissible signal-to-noise ratio and also the nonlinearity and limited dynamic range of the light source and detector make it desirable to employ digital signals. An obvious

choice is to use binary signals, for which the light source need only be switched on and off.

Each fibre in a cable has its own light source and detector. Some thirty or forty fibres, each already of wide bandwidth, combine to give a cable of very large transmission capacity. The line repeaters necessary for compensating the optical losses each consist of a detector and a light source with an electronic circuit between them for restoring the signal to its original shape and level.

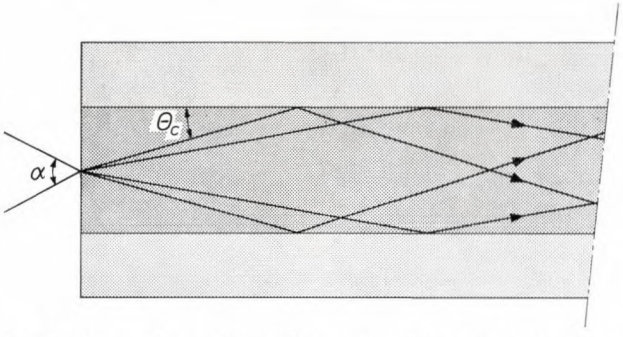


Fig. 2. Wave guidance by total internal reflection in a glass fibre. The maximum angle of incidence on the interface between the core and the cladding is θ_c. The magnitude of this critical angle is determined by the ratio of the refractive indices of the core and the cladding. The critical angle in turn determines the apex angle α of the acceptance cone of the glass fibre.

Physical parameters of the system

There are two groups of physical factors that set a limit to the system outlined above: those defining the power and those affecting the bandwidth of the system.

We shall look first at the requirements relating to the power. The shot noise resulting from the discrete nature of the photons begins to be of significance at low light intensities, and this, combined with the small probability of false detection that is normally allowed in a digital system, imposes a lower limit on the optical power at the detector end. At a modulation rate of 100 Mbit/s, for example, this lower limit has a value between 0.01 and 0.1 μW; at higher modulation rates the value is proportionately greater. If a repeater spacing of say 4 km were desired with fibres whose attenuation was 10 dB/km, it would be necessary to have an *effective* power of 1 mW available at the input. A power of 1 mW can be delivered by either an incoherently radiating diode or a laser diode. The question is however whether all of this power can be coupled *into* the fibre [6]. The area of the cross-section (diameter 50 μm) and the acceptance angle (24°) of the fibre are both of significance here. With these values the 1 mW of power will have to be delivered at a radiance of 285 W/cm²sr. At present this requirement can only be met with a laser; an LED can be considered solely for applications in which the fibre diameter can be much larger or the repeater spacing much less.

We shall then consider the limitations placed on the bandwidth in optical communications by means of a glass fibre. The modulation rate of 100 Mbit/s quoted above as an example can be achieved either with a GaAs laser or with an LED. Both laser and avalanche photodiode allow rates as high as several Gbit/s to be used [6].

Severe limitations can, however, be caused by the propagation in the glass fibre [7]. A short, sharply defined light pulse becomes lower and wider as it travels along the fibre. There are two reasons for this. The first is the colour dispersion of the glass; this has an effect because the spectral width of light from the source is finite. For GaAs-laser light whose combined spectral lines cover a spectral width of about 2 nm, there is a time-delay difference of 0.1 ns between the various components of the spectrum over a distance of 1 km. The light from an LED covers a width of about 40 nm and therefore causes time-delay differences and hence pulse spreading of 2.5 ns over one kilometre.

In addition, there are the delay differences between the various paths available to the light in a glass fibre. Let us take the extreme cases of light entering in the axial direction and light entering at the critical angle. The relative delay difference $\Delta t/t$ for these two directions of incidence is approximately equal to $\Delta n/n$. For the difference of 1% assumed earlier between the refractive indices of the core and the cladding, this gives a delay difference of 50 ns over a distance of one kilometre. If all angles of incidence between the two extremes are represented, an incoming pulse will be spread out over 50 ns during transmission over 1 km, giving a bandwidth of only 5 MHz for a repeater spacing of 4 km. Reduction of $\Delta n/n$ does not help since $\Delta n/n$ should not be much smaller than the 1% already referred to, or the acceptance angle of the fibre will become too small and large radiation losses will occur in small bends in the fibre. Reduction of the repeater spacing to obtain an acceptable bandwidth would require an impracticably small spacing. As we shall shortly see, however, there are other possibilities.

[1] See for example E. A. Aagaard, P. M. van den Avoort and F. W. de Vrijer, Philips tech. Rev. **36**, 85, 1976.
[2] See for example H. M. J. M. van Ass, P. Geittner, R. G. Gossink, D. Küppers and P. J. W. Severin, this issue, p. 182.
[3] See for example G. A. Acket, J. J. Daniele, W. Nijman, R. P. Tijburg and P. J. de Waard, this issue, p. 190.
[4] See for example L. J. M. Bollen, J. J. Goedbloed and E. T. J. M. Smeets, this issue, p. 205.
[5] See fig. 12 in the article by Van Ass, Geittner, Gossink, Küppers and Severin in this issue.
[6] See for example G. D. Khoe and L. J. Meuleman, this issue, p. 201.
[7] See for example J.-P. Hazan and L. Jacomme, this issue, p. 211.

The description of wave guidance in a glass fibre by means of total internal reflection, as given above, is in fact an over-simplification. For a proper description it is necessary to make use of the theory of propagation of electromagnetic waves. This theory shows that only certain modes of propagation are possible, which correspond to sharply defined directions of incidence of the light rays. Each of these modes has its own propagation velocity.

In large-diameter fibres the number of modes is so large that almost any direction of incidence between the axial direction and the critical angle is possible. The thinner the fibre, the smaller the number of modes that are possible until, finally, for an extremely small diameter only one mode will propagate.

The dispersion due to time-delay differences between the various light paths in a fibre in which many modes are possible (a 'multimode' fibre) is called mode dispersion. This dispersion, which can seriously restrict the bandwidth of an optical communication channel, as we have seen, can be avoided by using a fibre in which only one mode is possible, a 'single-mode' fibre. For a refractive-index difference of 1 % between the core and the cladding the core diameter to obtain a single-mode-fibre has to be smaller than about 3 μm. An objection to the single-mode fibre is that its small core diameter necessitates a light-source radiance that is several hundred times greater than the value already quoted for a fibre with a core diameter of 50 μm. A greater problem, however, is the coupling of the fibres; for a core diameter of only 3 μm this is by no means simple.

Another possible way of reducing delay differences between the modes in a multimode fibre consists in using a fibre in which the refractive index decreases gradually with the radius instead of abruptly. The light is then kept imprisoned in the fibre not by total reflection from a sharply defined interface but by being refracted gradually inwards in the outer layers of the fibre core. Theory indicates that the delay differences between the various light paths are least when the variation of the refractive index in the radial direction approximates to a parabola (*fig. 3*) [8]. The largest delay difference, between a light ray travelling along the axis and one that deviates the most from the axis, can then be expressed approximately by

$$\frac{\Delta t}{t} \approx \left(\frac{\Delta n}{n}\right)^2.$$

In the optimum case, once again for a refractive-index variation of 1 %, the delay difference between the various light paths is only 0.5 ns/km. Here the diameter of the fibre core can again be so large that there are relatively few problems, either with the optical power to be coupled into the fibre or the coupling of two fibres.

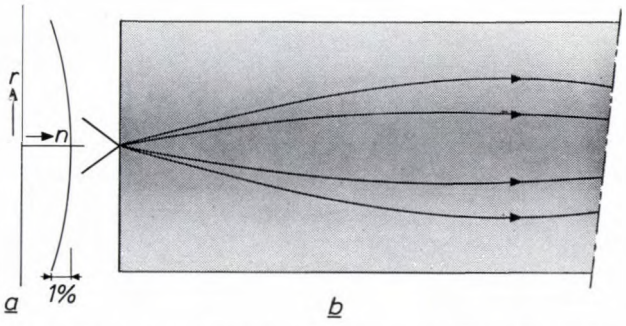


Fig. 3. The parabolic refractive-index profile. *a*) The variation of the refractive index as a function of the radius. The difference between the refractive index at the axis of the fibre and that at the outside of the fibre does not usually exceed 1 %. *b*) Light paths in a fibre with a parabolic (graded) refractive-index profile. Despite the difference in geometrical length between the various paths, the increase in the velocity of light towards the outside of the fibre ensures that the delay times are almost identical for all the paths.

Possibilities of an optical communication system

Now that the most important physical factors relating to optical communication on glass fibres have been summarized, we can look more generally at the maximum repeater spacing and the bandwidth available for combinations of the various light sources and fibres. In all cases it is assumed that the detector is an avalanche photodiode.

The performances of the various combinations have been calculated and are plotted in *fig. 4*. The calculations were based on an attenuation of 10 dB/km in the glass-fibre cable. Allowance has been made for the fact

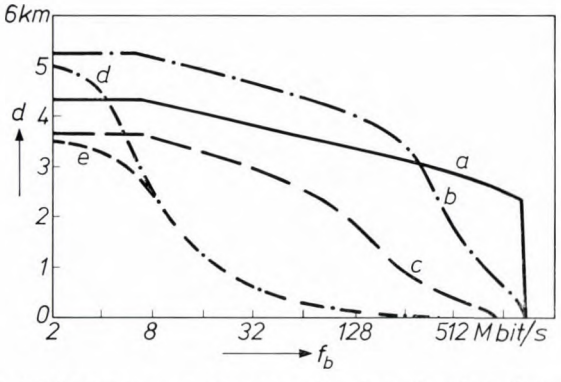


Fig. 4. The maximum repeater spacing d as a function of the bit rate f_b for combinations of different light sources and glass-fibre types. *a* laser with a single-mode fibre, *b* laser with a multimode fibre with a parabolic refractive-index profile. *c* the same fibre with a light-emitting diode (LED). *d* laser with a multimode fibre with a stepped refractive-index profile. *e* light-emitting diode with a multimode fibre, also with a stepped refractive index. A fibre attenuation of 10 dB/km is assumed in all cases.

[8] S. Kawakami and J.-I. Nishizawa, IEEE Trans. **MTT-16**, 814, 1968.

that the cable will be produced and laid in sections of limited length. Losses will inevitably occur where two lengths of cable are 'spliced'. The possibility of the cable being broken a number of times as a result of excavation work, etc. during the lifetime of the cable system — estimated at 30 years — has also been taken into account. The repair of these breaks will introduce additional coupling losses.

The figure shows clearly that the multimode fibre with a simple step in the refractive index will not give as great a maximum repeater spacing for increasing

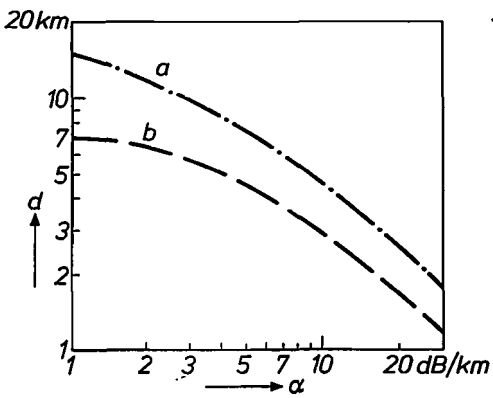


Fig. 5. The maximum repeater spacing d as a function of the cable attenuation α at a bit rate of 34 Mbit/s for a laser in combination with a multimode fibre with a parabolic refractive-index profile (curve *a*) and for a light-emitting diode in combination with a fibre of the same kind (curve *b*).

values of the bit rate as the fibre with the parabolically graded index or the single-mode fibre, and that a laser gives better results than a light-emitting diode. The combination of a light-emitting diode and a single-mode fibre is pointless since, as already observed, the radiance of an LED is not high enough to couple sufficient power into a fibre with such a thin core.

Fig. 5, finally, demonstrates the maximum repeater spacing as a function of the cable loss. The values were calculated for a modulation rate of 34 Mbit/s and the use of a glass fibre with a parabolic refractive-index profile, in combination with a laser or an LED.

In addition to telecommunication, where a large bandwidth and a large repeater spacing are of primary interest, there are many other applications in which optical communication on glass fibres can offer advantages for other reasons. These include applications in aircraft, where the low weight of the cables is an asset, and communication between installations some distance apart — e.g. instruments in hospitals or data-processing equipment in computer centres — in which it is an advantage not to have earth-potential differences to consider, and the insensitivity to electrical interference can also be invaluable. Yet another application of interest is in communication and control in high-voltage installations, where the possibility of bridging large voltage differences is of great value.

Summary. In the study of wideband transmission systems much attention is being given to an optical system in which information is conveyed in the form of light signals conducted along a glass fibre. If the refractive index of the glass fibre decreases towards the outside, light incident within a cone around the axis of the fibre will be contained within the fibre. In the last few years the optical losses per kilometre in such glass fibres have been reduced to a factor of about 2 per km. The light source may be either a GaAs laser diode or a light-emitting diode (LED) and a suitable detector is the Si avalanche photodiode. The bandwidth of a system of this kind depends chiefly on the differences in time delay for the various available light paths in the fibre. A large bandwidth can be obtained by making the fibre so thin that only one light path is possible or by giving the refractive index of the fibre a profile that decreases parabolically towards the outside. Systems with a transmission capacity of many hundreds of Mbit/s for a repeater spacing of many kilometres can be produced in this way. Advantages of an optical communication system of this kind are inexpensive and easily laid cables, its non-dependence on increasingly scarce copper, and its insensitivity to interference. The system is of interest not only for telecommunication but also for applications in aircraft, instrumentation, communication between equipment and in high-voltage installations.

The manufacture of glass fibres for optical communication

H. M. J. M. van Ass, P. Geittner, R. G. Gossink, D. Küppers and P. J. W. Severin

Introduction

From the very start the existence of the laser has been a stimulus to the idea of optical communication. One of the ideas put forward, a communication system in which signals are transmitted along optical waveguides by electrically modulated GaAs lasers, has led to studies of the potential of glass fibres in such an application. A fibre for optical communication must satisfy three requirements. First, as little radiation as possible should escape sideways (see p. 177); secondly, in the propagation inside the fibre the loss due to absorption and scattering should be as small as possible; thirdly, the shape of a light pulse should be preserved over large distances (low pulse dispersion). The first requirement is most easily met by making fibres that consist of a core (which conducts the light) and a cladding of a material of lower refractive index, giving total internal reflection at fairly low angles of incidence. If the refractive index is given a nearly parabolic, rather than stepped, variation in the radial direction, then in addition the lowest pulse dispersion is obtained [1].

In the wavelength range of interest here, the absorption of radiation in glass is mainly due to hydroxyl groups and ions of the transition metals (particularly Fe and Cu). It is therefore desirable to make the concentration of these impurities as low as possible. To limit the scattering every attempt must be made to prevent the presence of inhomogeneities, such as bubbles, crystallites and local variations in the refractive index.

Studies made in many countries have led to the conclusion that soft (multicomponent) glass of various kinds as well as doped silica glass can both be considered for the manufacture of fibres for optical communication. The techniques for making multicomponent-glass optical fibres have been based on conventional methods used in glass manufacture. In most of these techniques a double crucible has been used [2]. This article includes a description of a double-crucible method applied at Philips Research Laboratories, in which alkali-germanosilicate glasses are employed. If silica glass is to be used with its much higher melting point, there is much to be gained in using chemical vapour deposition (CVD). Such methods are well known from semiconductor technology [3]. At Philips Forschungslaboratorium Aachen a CVD method has been developed in which the chemical reaction is set up by a non-isothermal microwave plasma.

For multicomponent glass, the double-crucible process can be run continuously; this is an advantage for industrial production. In methods for processing silica glass, on the other hand, it is more easy to meet the requirements for high purity (low absorption). Of all the methods in which silica glass is the starting material, the plasma method seems to be the one most suitable for obtaining a parabolic index profile.

The double-crucible method; fibres of alkali-germanosilicate glass

In making a gradually varying index profile ('graded-index profile') in a glass fibre by the double-crucible method two kinds of multicomponent glass are used, of different refractive indices. The glass with the higher refractive index forms the core of the fibre and the other forms the outer portion or 'cladding'. An ion-exchange technique is used to give a smooth variation in the refractive index. It is clear that unless special measures are taken the index profile produced in this way by diffusion can only be a rough approximation to the desired parabolic profile.

A material containing monovalent ions is the most obvious choice for the ion exchange to give the desired index profile, since only monovalent ions diffuse rapidly enough. The exchange between Tl^+ and Na^+ or K^+ ions has usually been used previously, but there are problems with thallium: it is toxic, volatile at the fusion temperature of glass and has a tendency to assume valencies other than unity.

If two conventional types of glass are used as starting materials, an exchange between Na^+ and K^+ ions does not give a sufficiently large difference in refractive index. The desired difference can however be obtained by using alkali-germanosilicate glass. In addition to Na_2O or K_2O and SiO_2 this glass also contains GeO_2 as a main component, and small quantities of Al_2O_3 and CaO are added for stabilization.

Dr Ir H. M. J. M. van Ass, Dr Ir R. G. Gossink and Dr P. J. W. Severin are with Philips Research Laboratories, Eindhoven; Dr P. Geittner and Dr D. Küppers are with Philips GmbH Forschungslaboratorium Aachen, Aachen, West Germany.

To obtain the necessary purity, we make the raw material SiO_2 by hydrolysis of appropriate organic compounds [4]:

$$Si(OC_2H_5)_4 + 2H_2O \rightarrow SiO_2 + 4C_2H_5OH.$$

In this way products are obtained whose iron content is smaller than 0.1 ppm. The other raw materials (the alkali oxides in the form of carbonates) are commercially available in a sufficiently pure form.

The glass manufacture and the fibre-drawing processes are all carried out in the dust-free environment of a 'clean room'. The raw materials are put into a pure platinum crucible, and melted at a temperature of 1200-1500 °C in a resistance furnace, which has a silica lining (*fig. 1*). Good results have also been obtained with a water-cooled silica crucible in which the melt is inductively heated; the glass melt and the silica of the crucible wall do not then interact with one another.

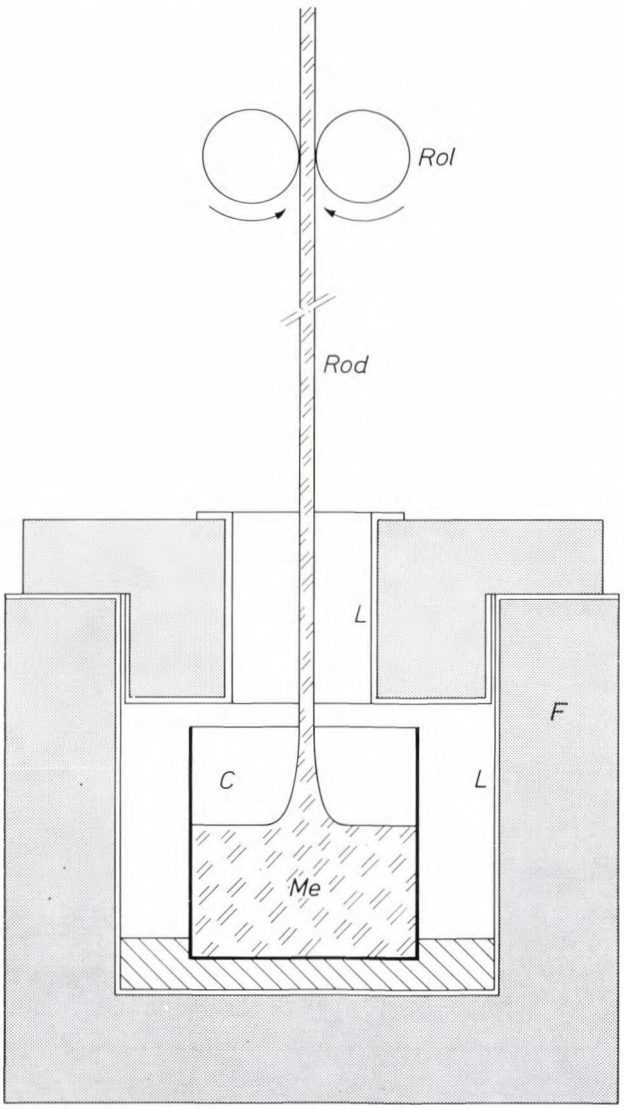


Fig. 1. Diagram of apparatus for melting glass and drawing the rods to be used in the double-crucible method (see fig. 2) of making glass fibre for optical communication. *F* resistance furnace, with silica lining *L*. *C* crucible of very pure platinum. *Me* melt of alkali-germanosilicate glass. *Rod* rod about 5 mm in diameter. *Rol* rollers to draw the rod upwards at constant speed.

During the melting process in which the alkali carbonates decompose and the various products react with one another to form the desired type of glass, carefully purified gas is blown through the melt to improve the homogeneity. In a later stage the melt is given another special temperature treatment (the 'fining'), to remove the gas bubbles arising mainly from the decomposed alkali carbonates.

In the next stage of the process the melt is brought to a temperature of 800 °C and rods, of circular cross-section and of diameter about 5 mm, are drawn from the melt at a rate of about 10 cm/min (fig. 1).

In this intermediate form the glass is suitable for storage and can easily be processed into fibres later. If desired, the glass can be submitted to a preselection at this stage, before the actual fibre manufacture.

The glass is drawn to a fibre with the aid of the double platinum crucible shown in *fig. 2*. Rods of the glass for the core are fed into the inner crucible, and rods of the glass for the cladding into the outer crucible. Care is taken to ensure that the melting takes place slowly and at a sufficiently high temperature to prevent inclusion of bubbles. The glasses flow out at about 800 °C. The desired ion exchange between the core and cladding glass takes place in the region between the outlets of the inner and outer crucibles, where the flow rate of the glass is very low. The choice of parameters such as the diffusion time and the temperature enables the radial distribution of the Na^+ and K^+ ions, and hence the shape of the index profile in the fibre, to be controlled within certain limits.

With equipment similar to that shown in fig. 7, a fibre of diameter 100 μm is drawn. Before the drawn fibre reaches the winding drum it is passed through a vessel containing a synthetic coating material, and then dried in a tubular furnace, so that it is covered with a thin protective coating.

Study of the optical losses; the refractive-index profile

We have used a differential method to determine the total optical loss (absorption plus scattering) of the fibres made in the above way as a function of the

[1] See the introductory article by K. Mouthaan in this issue, particularly p. 180.

[2] See for example K. Koizumi, Y. Ikeda, I. Kitano, M. Furukawa and T. Sumimoto, Appl. Optics **13**, 255, 1974.

[3] See for example:
F. P. Kapron, D. B. Keck and R. D. Maurer, Appl. Phys. Letters **17**, 423, 1970;
J. B. MacChesney, P. B. O'Connor, F. V. Di Marcello, J. R. Simpson and P. D. Lazay, Xth Int. Congress on Glass, Kyoto 1974, p. 6-40;
W. G. French, *ibid.*, p. 6-46.

[4] R. G. Gossink, H. A. M. Coenen, A. R. C. Engelfriet, M. L. Verheijke and J. C. Verplanke, Mat. Res. Bull. **10**, 35, 1975.

wavelength. The measured values are shown in *fig. 3*. For application in optical communication somewhat lower values are usually desirable.

In developing such a manufacturing method it is of course important to know whether the optical loss is perhaps mainly due to the intrinsic absorption of the glass selected, in which case the problems might be insuperable. This can best be investigated by measurements on the rods produced as an intermediate form. Since these are short this requires an extremely sensitive method.

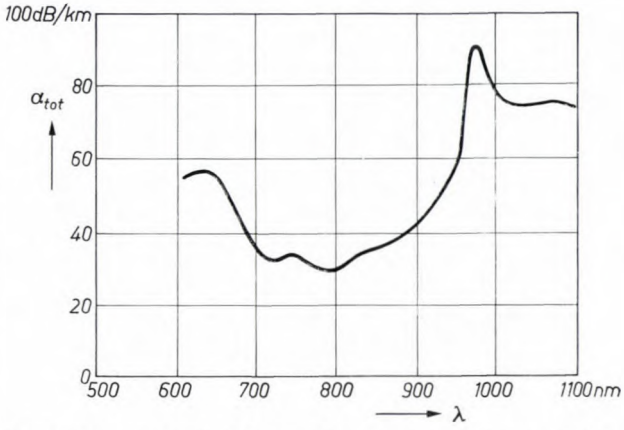


Fig. 3. The total optical loss α_{tot} for a fibre made by the double-crucible method as a function of the wavelength λ, determined from the difference in the absorption of a long fibre and a short one. The GaAs laser to be used emits radiation in the wavelength region between 800 and 900 nm. The peak at 970 nm is due to hydroxyl groups.

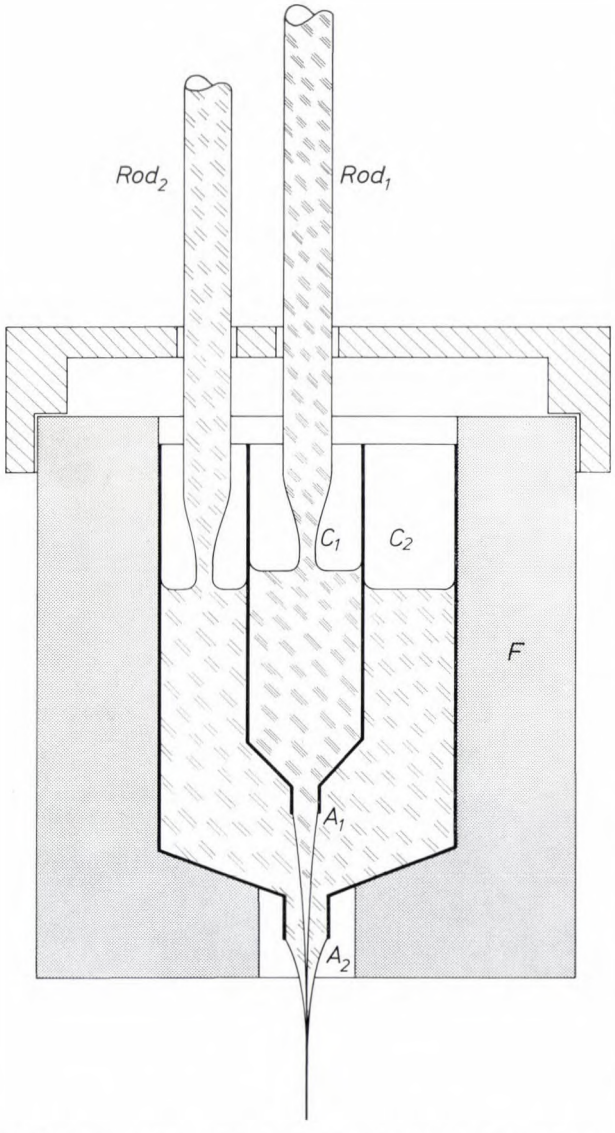


Fig. 2. The double crucible for making glass fibre. C_1 and C_2 two concentric crucibles of pure platinum. Rod_1 and Rod_2 rods made in the apparatus shown in fig. 1; Rod_1 consists of alkali-germanosilicate glass with a high Na^+-ion content, and hence a high refractive index, Rod_2 of the same glass but with a high K^+-ion content, and a low refractive index. A_1 outlet of the inner crucible C_1, A_2 outlet of the outer crucible C_2. In the region between A_1 and A_2 there is an exchange between the Na^+ and K^+ ions, which gives a smooth variation for the refractive-index profile in the resulting fibre. F resistance furnace.

We have used a *calorimetric* method for this purpose [5]. This is based on the increase in temperature caused by incident light absorbed in the rod. The increase in temperature is measured by thermocouples pressed against the surface of the rod. With this arrangement it has been found that the absorption of the glass rods is less than 10 dB/km at a wavelength of 750 nm, indicating that the choice of germanosilicate glass is not incompatible with the desired objective.

The measurements we have just mentioned are carried out in the arrangement shown in *fig. 4*. At the equilibrium temperature the absorbed light energy is equal to the losses by radiation and conduction. These losses are difficult to determine accurately, however. If, in addition to the increase in temperature, the relaxation time is measured with which the temperature of the surface of the rod decreases to the ambient value when the incident light is removed, the losses can be eliminated from these two quantities.

Since the rods to be examined have a low absorption coefficient and the temperature will only increase slightly, and since for small heat losses — the increase in temperature is then the largest possible — the relaxation times become very long, our test equipment is arranged in such a way that the temperature fluctuations are as small as possible over the longest possible time.

The refractive-index profile of the fibres under test can be measured with a Michelson interferometer. *Fig. 5* shows the differences in optical pathlength made visible by a Michelson interferometer for a disc, 50 μm thick, sawn from a fibre embedded in a capillary tube. The outer region with parallel fringes corresponds to the constant refractive index of the capillary tube. The curved fringes in the central region represent the change in the refractive index in the fibre. They show that the approximation to the desired variation is quite good.

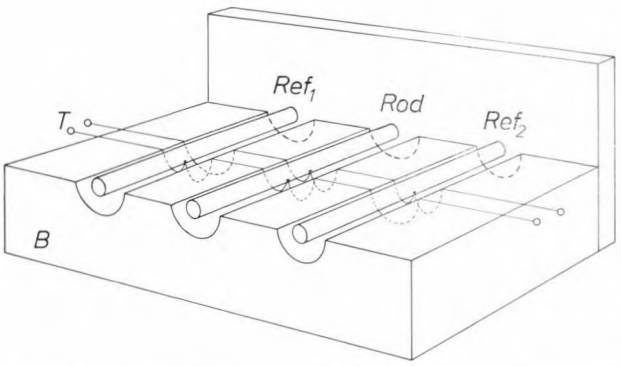


Fig. 4. Diagram of a calorimetric test arrangement for determining the absorption in short glass rods. The rod to be measured *Rod* and two reference rods Ref_1 and Ref_2 are located in a copper block *B* (the upper half is not shown). The surface temperature of these rods is measured by copper-constantan-copper thermocouples *T*, which make a light spring contact with the rods. The arrangement is duplicated to allow good contact with the rods to be verified experimentally. This arrangement is completely enclosed in an evacuated bell-jar, which is located in a water bath held at constant temperature. Temperature fluctuations can be limited to 10^{-5} K in this way. The quantities measured are the increase in temperature, which is due to the absorption in the rods of incident light in the wavelength range 400 to 800 nm from a Kr-ion laser, and the relaxation time with which the temperature decreases to the ambient value when the incident light is removed. From these two quantities the absorption can be determined to an accuracy of 0.02 dB/km for an incident power of 100 mW.

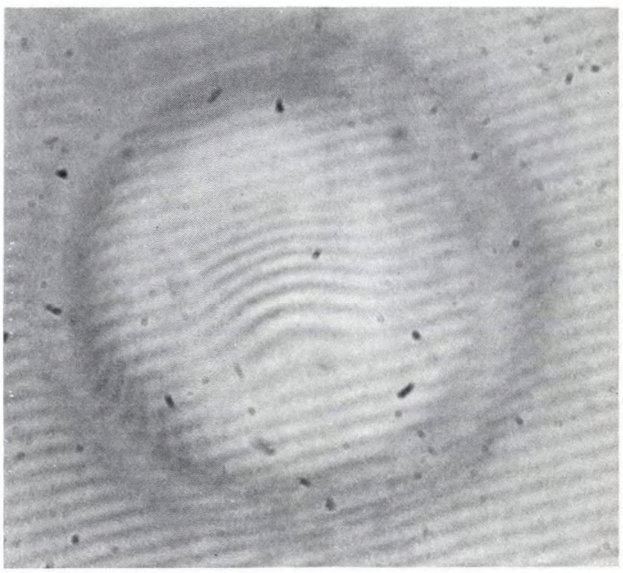


Fig. 5. Interference pattern obtained with a Michelson interferometer for a disc sawn from a glass fibre. The fibre, with an outer diameter of 100 μm, was made by the double-crucible method; the disc, 50 μm thick, is embedded in a capillary tube. The parallel fringes indicate the region of constant refractive index, the curved ones indicate the region of changing refractive index, in which the shape of the curved fringes gives a general picture of the variation of the refractive index.

The plasma method

Quite a different procedure is used for making silica-glass fibres. There are three stages to the process. During the first stage gaseous $SiCl_4$ and $GeCl_4$ are allowed to react with oxygen in a tube of pure silica glass:

$$SiCl_4 + O_2 \rightarrow SiO_2 + 2Cl_2,$$
$$GeCl_4 + O_2 \rightarrow GeO_2 + 2Cl_2.$$

The oxides produced in these reactions are deposited on the inner surface of the tube (which will later become the cladding) while a local reaction zone passes over it. The $GeCl_4$ content of the gas mixture is increased in steps to produce a gradual increase in the refractive index.

In the method that has previously been used [3] the reaction zone is a small region at a temperature of 1400-1600 °C obtained by heating with oxy-hydrogen burners. 'Glass-soot' produced in this zone is deposited on the tube wall and melts there to form a coating layer. To obtain as even a layer as possible the tube is rotated.

In the method we shall now describe, this local hot zone is replaced by a plasma, produced by a microwave resonator that moves to and fro. Advantages of this method are that the deposition zone is small and that the speed of the resonator can be very high. This means that very even and thin layers can be obtained at each pass of the reaction zone, and a good approximation to the desired index profile is therefore possible [6]. In this method, unlike the thermal method, there is no formation of 'soot', so that the tube does not have to be rotated. In addition, the process is not operated at atmospheric pressure, but at about 1300 Pa (10 torr).

Fig. 6 is a diagram of our experimental arrangement. The highly purified gases $SiCl_4$, $GeCl_4$ and O_2 are passed through the silica tube *Tu* in the correct quantities, regulated by the accurately adjustable flow-control valves *R*. Under the conditions that we have chosen there is a complete conversion of SiC_4 and $GeCl_4$ into silica glass doped with germanium dioxide. The chlorine gas liberated in this reaction is removed by a sorption pump *Pu* which is filled with zeolite and cooled with liquid nitrogen. The tube *Tu* is located in the furnace *F*, which is held at a temperature of about 1000 °C during the process; this prevents the formation of cracks resulting from stresses in the glass. *P* is the plasma excited in the microwave resonator *M*, which moves rapidly backwards and forwards.

The dosage of the gases and the speed with which the resonator moves along the tube determine the thickness

[5] See H. N. Daglish and J. C. North, IXth Int. Congress on Glass, Versailles 1971, Part 1, p. 769.
[6] P. Geittner, D. Küppers and H. Lydtin, Appl. Phys. Letters **28**, 645, 1976.

of the layer produced, and hence the number of layers that will eventually form the core. Control of the $GeCl_4$ supply enables the index profile of the glass to be varied within wide limits. Since the complete process takes place in the silica tube, which is sealed off from the ambient atmosphere, it is easier to meet the requirements for high purity than in the double-crucible method.

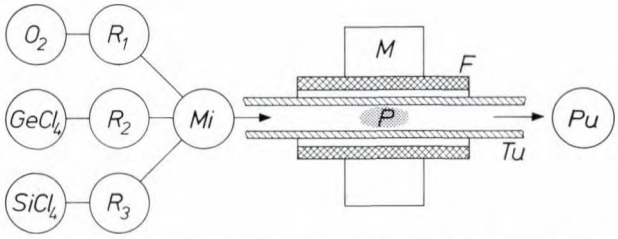


Fig. 6. Diagram of the arrangement for the first stage in the plasma method. *Tu* silica tube, which will later become the cladding of the glass fibre. *F* resistance furnace. *M* microwave resonator for generating the plasma *P*; the resonator is moved rapidly back and forth. R_1, R_2 and R_3 flow-control valves for admitting accurately known quantities of the gases $SiCl_4$, $GeCl_4$ and O_2 to the reaction region. *Mi* mixer. *Pu* sorption pump, filled with zeolite and cooled by liquid nitrogen.

At each pass of the plasma a layer of silica glass doped with germanium dioxide is deposited on the wall of the tube *Tu*. Any desired index profile — over a wide range — can be obtained by varying the supply of $GeCl_4$.

After the successive layers of glass have been formed in the tube as described, the second stage of the process follows. A ring burner at a temperature of about 2000 °C is passed along the tube a number of times. Since this temperature is above the softening point of silica glass, this treatment produces a gradual reduction in the tube diameter because of the presence of surface tensions. Eventually the tube collapses to form a rod with core and cladding, the 'preform'.

In the third and final stage of the process this preform is drawn in the equipment shown in *fig. 7* to form a fibre with an outer diameter of about 100 μm and a core diameter of about 50 μm. In this stage the preform is heated to about 2000 °C again, either in a graphite furnace or in a ring burner. The equipment shown in fig. 7 includes ring burners.

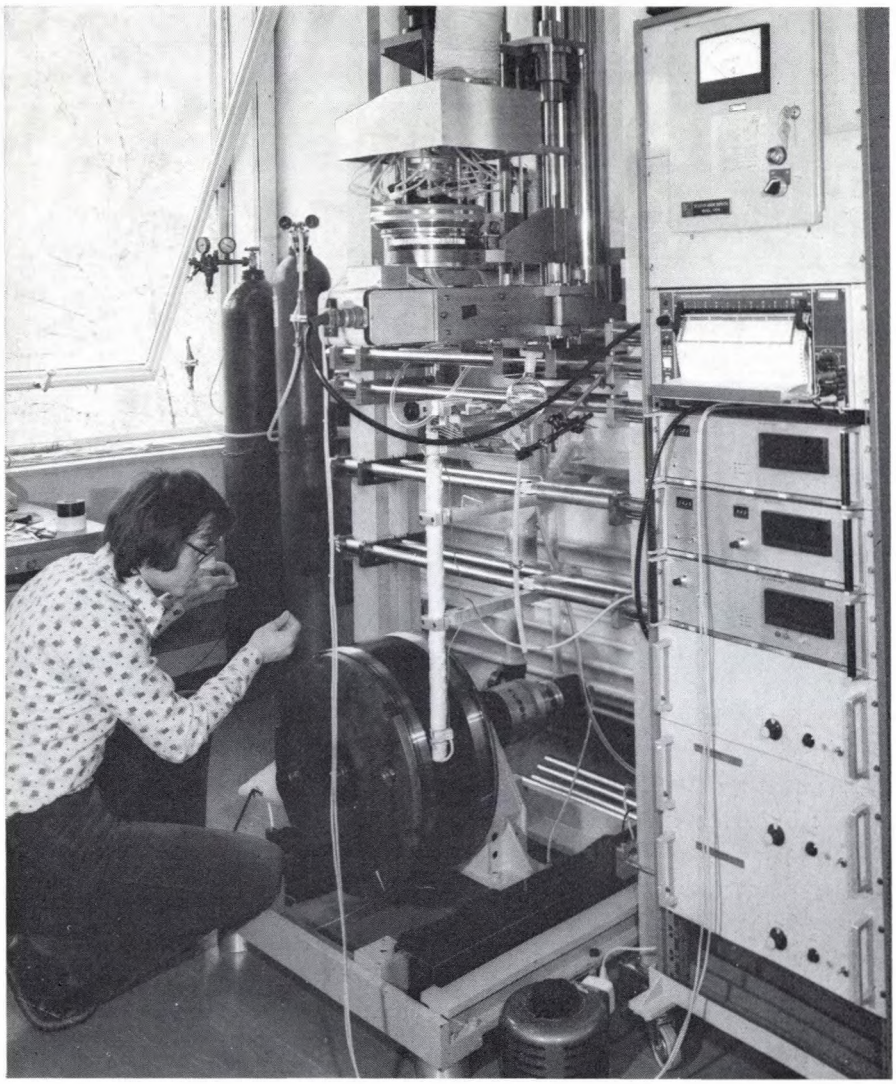


Fig. 7. View of the equipment for drawing silica-glass fibres [*]. Extraction equipment can be seen at the top, and beneath this the supply tubes for the burners. The upper end of the preform to be drawn can just be seen to the left of the thick extraction pipe. After the fibre has been drawn it passes through a container of synthetic coating material and through a tubular furnace. This provides a thin protective coating for the fibre, which is then wound on to the drum. The box above the furnace contains optical equipment for measuring the diameter of the fibre.

Results of investigations and measurements

We started our examination of the process with a study of the deposition of the glass in a stationary resonator. *Fig. 8* illustrates the tube gradually closing up under these conditions. The gas mixture here flowed from left to right. The deposition zone is about 10 mm long. *In fig. 9* the quantities of glass deposited and the concentration of GeO_2 it contains are both plotted in arbitrary units along the axial direction of the deposition zone. It can be seen that the concentration of GeO_2 is not constant in the axial direction. This leads to the conclusion that the deposition kinetics of SiO_2 and GeO_2 are not the same [7].

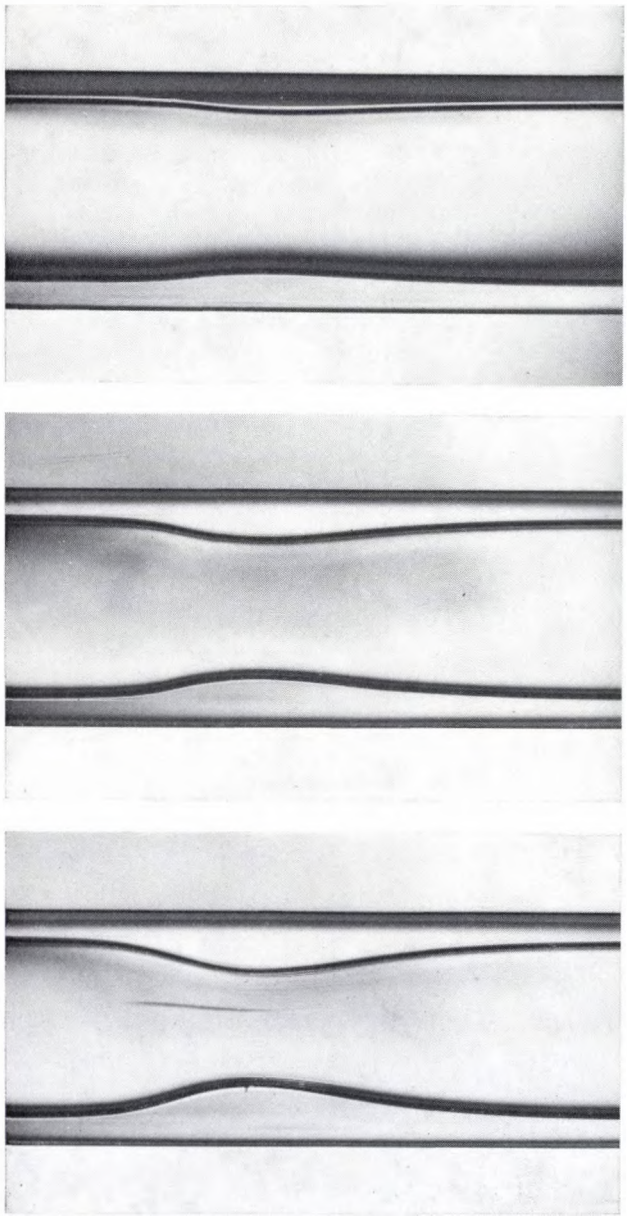


Fig. 8. Longitudinal cross-sections of a silica-glass tube after deposition of pure silica glass on the inner surface for 10, 20 and 30 minutes by the plasma method with the resonator stationary. The gases $SiCl_4$ and O_2 flowed from left to right. Magnification 4×.

In deposition with the resonator moving to and fro this variation in the content of GeO_2 leads to fluctuations in the refractive index in each separate layer, as *fig. 10a* shows. This photograph shows the interference pattern obtained with a Michelson interferometer for a disc 250 μm thick, sawn from a preform. The core of the preform was made from 45 layers, with a constant supply of $GeCl_4$. The step at the boundary layer corresponds to the higher refractive index resulting from the supply of $GeCl_4$. It can be seen that there is a 'wavy' pattern for the variation inside the core, even though the gas composition was constant during the deposition of the glass layers. This is connected with the

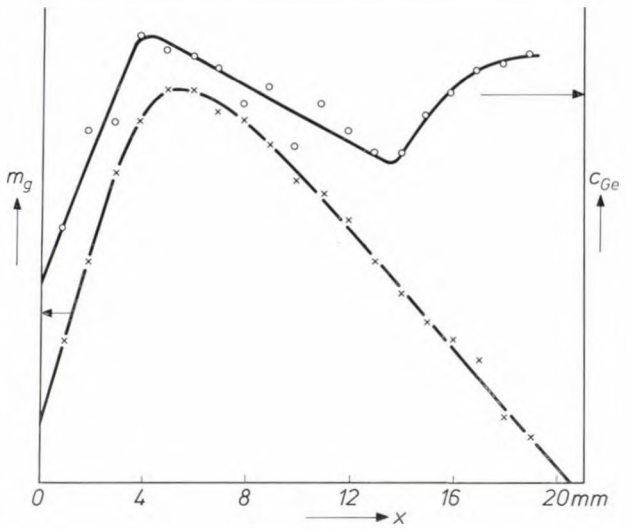


Fig. 9. The quantities m_g of glass deposited (lower curve) and the concentration c_{Ge} of GeO_2 in it, in arbitrary units and plotted against the axial distance x in a deposition zone as in fig. 8.

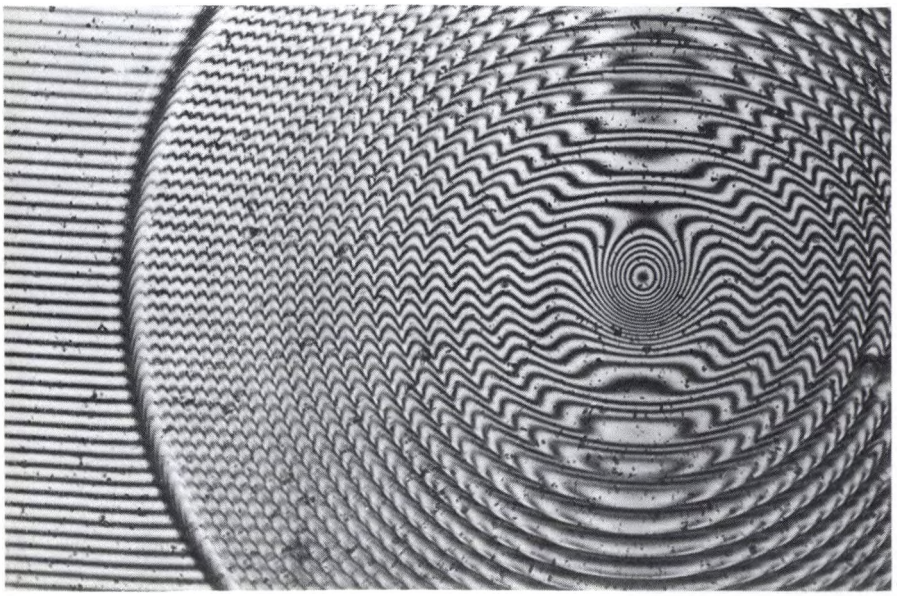
difference between the deposition kinetics of SiO_2 and GeO_2.

Fig. 10*b* gives a similar interference pattern, but now for a preform in which the (about 2000) layers were deposited with a $GeCl_4$ supply that increased linearly with time. The large number of layers has given a good approximation to the desired parabolic variation, though there is a dip at the centre. *Fig. 11* shows the variation in the germanium concentration across the core diameter, measured with a microprobe. Here again the dip is clearly visible.

The low germanium concentration at the centre is a result of the evaporation of GeO_2 from the region close to the surface, when the tube collapses to produce a

[7] D. Küppers, J. Koenings and H. Wilson, J. Electrochem. Soc. **123**, 1079, 1976.

[*] The fibre-drawing equipment was designed and constructed by W. C. P. M. Meerman (Philips Research Laboratories, Eindhoven); the burners used for silica glass were constructed by K. D. Zimmermann (Philips GmbH Forschungslaboratorium Aachen).



a



b

Fig. 10. *a*) Michelson-interferometer pattern for a 250 μm thick disc, sawn from a preform made with a constant supply of $GeCl_4$. The core (diameter about 3 mm) consists of 45 layers. *b*) As (*a*) but with a linearly increasing supply of $GeCl_4$. The number of layers here is 2000.

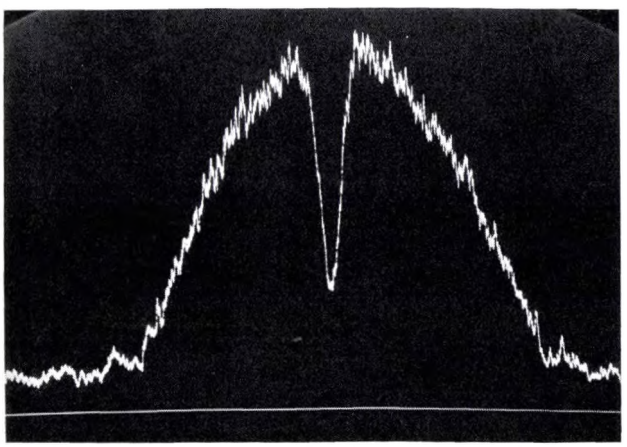


Fig. 11. Microprobe measurement of the variation in the germanium concentration across the diameter of a preform.

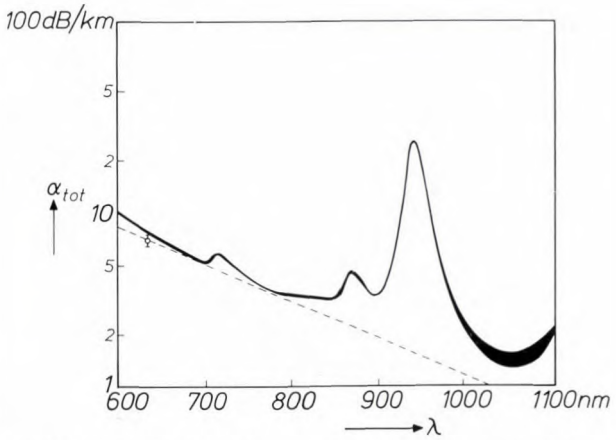


Fig. 12. The total optical loss α_{tot} of a fibre made by the plasma method as a function of the wavelength λ, and also the scattering loss, determined at a wavelength of 633 nm and extrapolated (dashed line). In both cases the measurements were made with the optical bridge shown schematically in fig. 13.

[8] Measurements of this type are described in the article by J.-P. Hazan and L. Jacomme in this issue, p. 211.

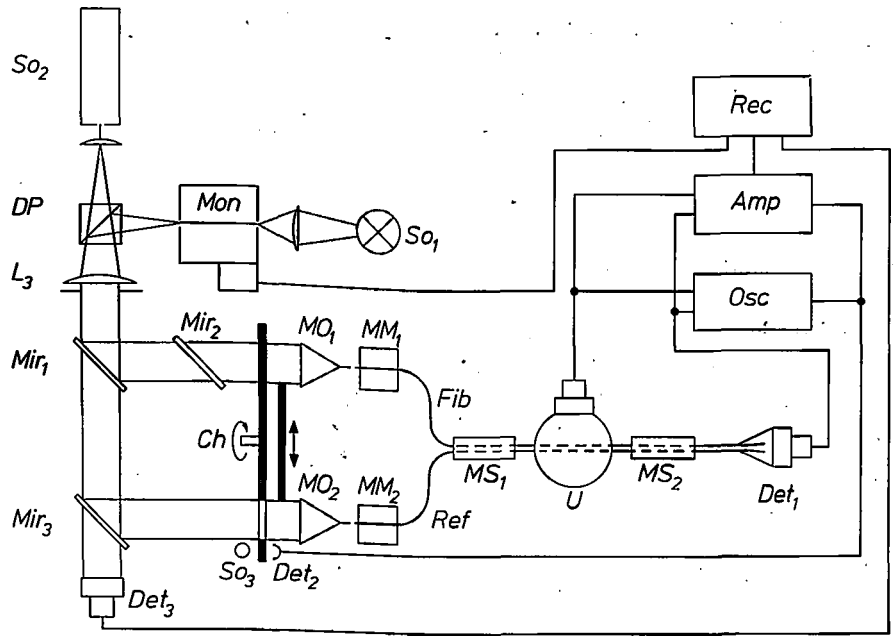


Fig. 13. Diagram of a bridge for determining the total optical loss and light scattering of a glass fibre. Two light beams of equal intensity are produced with the aid of semitransparent mirrors Mir_1, Mir_2 and Mir_3. One beam is transmitted via the microscope objective MO_1 and the micromanipulator MM_1 to the fibre *Fib* under test, and the other via MO_2 and MM_2 to the reference fibre *Ref*. This reference fibre is so short that the attenuation in it is negligible. The light beam is offered alternately to the test fibre and the reference fibre by the action of the chopper *Ch*. The transmitted light from both fibres is measured by the photodiode Det_1. The diode signal is measured with the phase-sensitive amplifier *Amp*. The phase-reference signal is supplied by the detector Det_2 and the light source So_3; the light from So_3 is also chopped periodically by *Ch*. The scattered light is measured with the Ulbricht sphere *U*. The light can originate from the He-Ne laser So_2 for measuring the scattering at a discrete wavelength or from the combination of the tungsten-band lamp So_1 and the monochromator *Mon*, for measuring the total optical losses as a function of the wavelength (the intensity of the monochromator is insufficient for measuring the scattering losses with the Ulbricht sphere). Det_3 reference photodiode. *DP* double prism. MS_1 and MS_2 modestrippers that collect the light escaping sideways. *Osc* oscilloscope. *Rec* recording equipment.

preform. Fortunately, such a deviation from the parabolic curve has no serious consequences: measurements have shown that the pulse dispersion is only 0.8 ns/km[8].

Results of measurements on the fibres are shown in *fig. 12*. The measurements were performed with a differential bridge, as shown schematically in *fig. 13*. With this bridge both the total optical loss and the part of the loss due to scattering can be measured. It can be seen that the total optical losses in the wavelength region of interest lie between 3 and 5 dB/km (with a minimum of 1.4 dB/km at 1.06 μm), and hence well below the value normally considered acceptable in practice for optical communication.

Summary. Description of two methods for the manufacture of glass fibres for optical communication that have a low optical loss and a low pulse dispersion. The fibres consist of a cladding around a core of higher refractive index, which has a nearly parabolic variation across the core diameter.

In the double-crucible method rods of alkali-germanosilicate glass are fed into two concentric crucibles; the rods differ in their content of Na^+ and K^+ ions, and therefore in refractive index. In the region in which the two streams of molten glass merge to form the fibre there is an exchange between the Na^+ and K^+ ions; this ion exchange is such as to produce a smooth, approximately parabolic variation in the refractive index.

In the plasma method silica glass is deposited layer by layer on the inside of a silica tube. The deposited glass is doped with germanium dioxide to give the desired variation in refractive index. The deposition in layers is carried out with the aid of a microwave resonator that moves to and fro in the axial direction. This last method gives fibres with a total optical loss of 3 to 5 dB/km and a pulse dispersion of 0.8 ns/km.

Semiconductor lasers for optical communication

G. A. Acket, J. J. Daniele, W. Nijman, R. P. Tijburg and P. J. de Waard

Introduction

As has already been said in the introductory article[1], the semiconductor laser is an attractive light source for optical communication: its dimensions are small (and reasonably well matched to those of a cross-section of an optical glass fibre), it uses little energy and is relatively easy to modulate (directly by means of the drive current, so that there is no need to use expensive electro-optical elements). The principle of the semiconductor laser, of gallium arsenide, has been known since 1962 [2], but the early prototypes needed a very high drive current to produce laser action. This made continuous operation at room temperature impossible. Since then, however, many improvements have been introduced, mainly through the use of a multilayer structure, which has now made continuous laser action at room temperature possible. In this article we shall discuss our work on a laser with a double heterojunction, which requires only a low drive current. We shall begin by giving a short account of the operation of semiconductor lasers in general. After dealing with the structure and operation of the multilayer laser for optical communication, we shall examine the choice of device parameters and show how these lasers are made by means of advanced crystal-growth methods. The article concludes with a discussion of the effects that determine the life of the laser.

Laser action in semiconductors

The 'classical' light sources, such as incandescent lamps and discharge lamps, emit incoherent radiation: the different surface elements behave like independent radiators, and the waves transmitted through one small surface element possess little or no phase correlation with time. This is because all these light sources are based on spontaneous emission of photons by excited atoms or ions. Since the beginning of the sixties coherent light sources have also become available: in these sources there is a well defined phase correlation over the radiating surface, and there is also phase coherence within reasonably long time intervals. These light sources, the lasers, are based on *stimulated* emission. Spontaneous and stimulated emission may be compared with one another by reference to the diagrams given in *fig. 1*, where line (*1*) represents the energy level of an electron in its ground state and line (*2*) the electron energy level in the excited state. If the electron is in the ground state (*a*) it can be raised to state (*2*) by the absorption of a photon; an electron in state (*2*) can return to state (*1*) accompanied by the emission of a photon either *spontaneously*, i.e. without interaction with any radiation field (*b*), or on *stimulation* by photons already present (*c*). In case (*c*) the emitted photon is in the same state — in classical terms: it is in phase with — the photon that caused the stimulated emission. In this mechanism large numbers of atoms can emit photons in phase and thus send out coherent radiation; this is the laser effect. The stimulated emission will only exceed the absorption when the number of the electrons in the excited state is greater than that in the ground state (population inversion). In addition the net surplus of stimulated emission must be appreciably greater than the spontaneous emission, in other words enough photons of a particular kind (light waves) must be present to be able to cause sufficiently strong stimulated emission. In most lasers this is accomplished by fitting a set of mirrors that reflect the

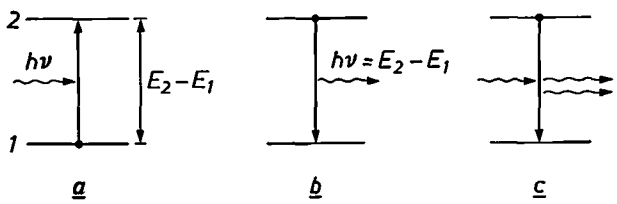


Fig. 1. Schematic representation of electron transition on absorption (*a*), spontaneous emission (*b*) and stimulated emission (*c*). In the first case a radiation quantum whose energy $h\nu$ is equal to the energy difference $E_2 - E_1$ of the energy levels *1* and *2* causes an electron to move from the lower to the higher level. In the two other cases the transition takes place in the other direction, accompanied by the emission of a quantum of energy $E_2 - E_1$. In stimulated emission this results from the incidence of a quantum of the same energy. In stimulated emission the incident and emergent waves are in phase.

Dr G. A. Acket, Ir W. Nijman, R. P. Tijburg and Ir P. J. de Waard are with Philips Research Laboratories, Eindhoven; Dr J. J. Daniele is with Philips Laboratories, Briarcliff Manor, N.Y., U.S.A.

light waves produced by stimulated emission back to the active laser medium, so that a standing wave is set up between the mirrors (optical resonator). When one of the mirrors is made partially transparent, some of the radiation can pass through and a laser beam is obtained.

Gas lasers, dye lasers, semiconductor lasers and other solid-state lasers are all based on the above principles. In semiconductors the electrons do not occupy discrete levels, as in the foregoing example, but energy *bands*. In an insulator the valence band is completely filled and the conduction band is completely empty. In a semiconductor there are a number of electrons in the lower levels of the conduction band and a number of holes in the upper levels of the valence band. These numbers can be increased, and thus part of the required population inversion can be achieved, by using a *P-N* junction diode, forward-biased so that there is a flow of current, instead of a homogeneous semiconductor. On each side of the junction there are then both additional electrons and additional holes, and at a sufficiently high current a state of population inversion arises.

The optical resonator is fairly easily produced in semiconductor materials by making use of the fact that crystals have preferred cleavage planes, which of course are very accurately parallel. In the material that we use, GaAs, the difference between its refractive index and that of air produces a sufficiently large reflection, and therefore no mirrors are required.

The choice of GaAs as a semiconductor is due to the fact that, unlike the more familiar semiconductors like silicon and germanium, gallium arsenide is a 'direct' semiconductor. In a material of this type direct recombination is possible between holes and electrons, accompanied by the emission of a photon, which implies a high probability of stimulated emission.

The injection laser with double heterojunction

The injection laser we have developed consists of a number of epitaxially grown layers of the mixed crystal $Al_xGa_{1-x}As$ ($0 \leqslant x < 1$) on a GaAs substrate. A schematic illustration can be seen in *fig. 2*. The current I passes successively through three layers of *P*-type conduction and finally through a layer of *N*-type conduction. At the interface of layers *2* and *1* electrons are injected into layer *2*; this layer consists of pure GaAs ($x = 0$).

Fig. 3 shows a simplified energy-band diagram of the system. The band gap E_g in layers *1* and *3* is greater than that in layer *2*. This is compensated between *1* and *2* by a step ΔE in the valence band and between *2* and *3* by an equally large step in the conduction band. The magnitude of the step is about $1.4x$ eV. These steps form potential barriers for holes (between *1* and *2*) and electrons (between *2* and *3*) which are injected into the GaAs layer, from which they are no longer able to escape. Recombination takes place accompanied by the emission of radiation with a quantum energy of about 1.4 eV, corresponding to a wavelength of 885 nm

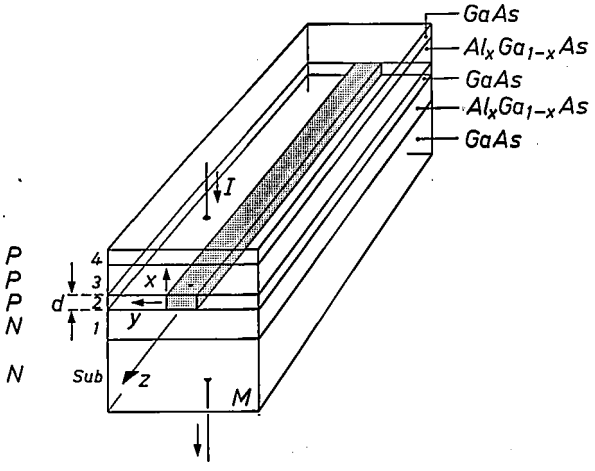


Fig. 2. Schematic representation of a GaAs laser with a double heterojunction. The laser action takes place in a narrow zone in the *P*-GaAs layer *2* (shaded area). The planes *M* are cleavage planes of the crystal which act as mirrors. The radiation is emitted in the *z*-direction. The laser action is brought about by a direct current which is passed through the whole structure in the *x*-direction. In production, the *N*-GaAs layer *Sub* acts as the substrate on which the layers *1* to *4* are epitaxially deposited. The thickness *d* of the active layer is only about 0.3 μm; the thickness of the other layers varies from 1 to 2 μm.

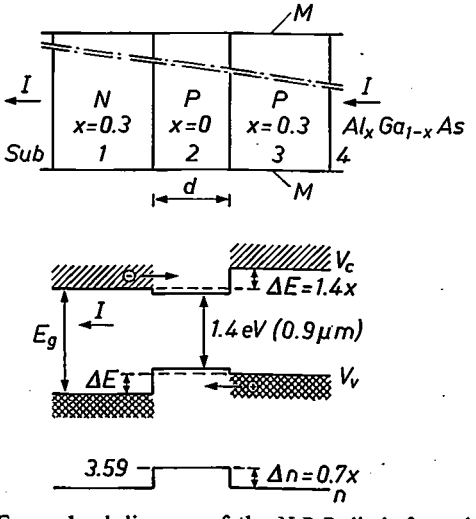


Fig. 3. Energy-level diagram of the *N-P-P* diode formed by the layers *1*, *2* and *3* (rotated by 90° with respect to fig. 2). When a current *I* flows in the direction indicated, electrons and holes are injected into layer *2*, and are prevented from escaping by the presence of the step ΔE in the height of the appropriate band, caused by the presence of a certain proportion of aluminium in layer *1* and in layer *3* ($\Delta E \approx 1.4x$ eV). The aluminium also causes a step Δn of magnitude $0.7x$ in the refractive index.

[1] K. Mouthaan, this issue, p. 178.
[2] See for example R. N. Hall, G. E. Fenner, J. D. Kingsley, T. J. Soltys and R. O. Carlson, Phys. Rev. Letters 9, 366, 1962.

(i.e. the near infrared). To guarantee the trapping effect of the potential barriers the step ΔE must be large compared with the thermal energy kT of the charge carriers, which means in practice that x must be greater than 0.1.

The concentration of charge carriers n_{th} at which inversion followed by amplification takes place in the active GaAs layer is of the order of 1 to 2×10^{18} cm^{-3}. From this we can estimate the minimum current density j_{th} at which stimulated emission will occur in any given laser. The concentration n_{th}, the elementary charge e, the current density j_{th}, the lifetime τ of an electron for direct recombination and the thickness d of the GaAs layer, which is filled almost homogeneously by the electrons and holes, are related by:

$$n_{th}e = j_{th}\tau/d. \qquad (1a)$$

Since τ is about 5 ns [3], it follows that

$$j_{th}/d \approx 5 \text{ kA/cm}^2\,\mu\text{m}. \qquad (1b)$$

The threshold current of a laser with 'suitable' dimensions is 100 to 200 mA, the operating voltage 2 to 3 V.

In the laser illustrated in fig. 2 the light is not confined in the x-direction between outer surfaces of the crystal but between the interfaces of layer *2* and the adjoining layers. Since the fraction x in these layers is about 0.3, their refractive index differs from that of layer *2* by about 0.2, the difference in refractive index being about $0.7x$. This results in such a large critical angle for total internal reflection that a sufficiently large part of the radiation remains within the active zone [4]. We shall return below to the consequences of the fact that the thickness of the layer in which the light is confined here is comparable with the wavelength of the light. The laser is bounded in the y-direction by making layers *3* and *4*, except for a central strip, into insulators by means of proton bombardment.

The injection laser as a waveguide; choice of parameters

A laser of the type outlined above may be regarded as a dielectric waveguide. Assuming that the electromagnetic field is independent of the coordinate y (see fig. 2), the Maxwell's equations that describe this field can be solved by a standard method [5]. It is then found that two types of field configuration (modes) can occur; the dominant one being a mode with the electric field vector in the plane of the junction and normal to the direction of propagation. This is because the reflection coefficient of the mirrors is greater for this mode than for the other mode [6]. In the direction normal to the active layer the electromagnetic field has a cosine distribution; outside the layer it decreases exponentially. A simple quantity used for characterizing this mode is the half-width of the square of the electric field-strength, to which the intensity is proportional. *Fig. 4* shows how this width depends on the thickness d of the active layer for various values of the Al fraction x. As can be seen, the curves have a minimum whose depth increases with the value of x. If a large d is chosen (as in the right-hand half of the figure) modes of higher order can then occur at large values of x, i.e. for large steps in the refractive index. The electromagnetic field then has more maxima. For practical reasons these modes are often undesirable. If x is small, the width of the field also increases considerably to the left of the minimum. The confining effect of the waveguide is then lost and a decreasing part Γ of the total

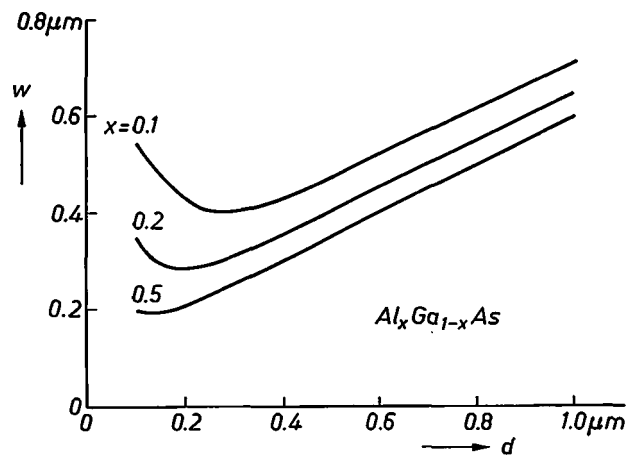


Fig. 4. The half-width w of the intensity distribution of the radiation, in the direction perpendicular to the active layer (see fig. 2), as a function of the thickness d of this layer for three values of the aluminium fraction x in the adjacent layers (calculated).

power in the wave remains within the active layer. The average amplification of the waves by stimulated emission in the active layer therefore decreases by the factor Γ, with the result that the density of the injected current has to be higher than the value of $5d$ kA/cm^2 found earlier (d is in μm). This is illustrated in *fig. 5*, where the threshold current is plotted as a function of d for a large number of diodes with $x = 0.3$. (Most of these diodes were made from slices obtained by optimizing the production process; see the next section.)

When the electromagnetic field in the laser is known, it is possible to calculate the spatial intensity distribution of the emergent radiation [7]. Unlike the situation in gas lasers, for example, where the exit aperture is large compared with the wavelength and the exit beam is accurately parallel, here the aperture of the exit beam can be many tens of degrees; see *fig. 6*. The fact that this quantity is also to some extent under

control makes it easier to design the best laser for a given type of glass fibre.

The above theory applies equally of course to the confinement of the electromagnetic field in the plane of the active layer, i.e. in the y-direction. Here the required step in the refractive index is brought about by means of the proton bombardment mentioned earlier, leading to a Δn of about 0.01. (As we noted, the protons do not penetrate into the active layer itself. The change in the refractive index is an effect connected with the current distribution in the active layer.) Since Δn is much smaller here than in the x-direction, the width of the active strip in the y-direction can be greater without modes of higher order occurring (see fig. 4). We use a width of about 5 μm. Because of this greater width the aperture of the beam in the y,z-plane is much smaller than that in the x,z-plane.

Finally, a few words about the oscillation condition that must be satisfied for the laser to operate, since this condition reveals the length that has to be chosen and indicates how a low drive current can be achieved. The oscillation condition can immediately be written after determining what happens to a wave leaving one of the two mirrors; see *fig. 7*. A wave of intensity I_0 grows on its way to the other mirror by a factor $\exp(\Gamma G - \alpha)l$. Here G is the gain in the active layer (per unit pathlength) — it depends linearly on the current density — α is the attenuation of the wave caused by the presence of free charge carriers, and l the distance travelled. Upon arrival at the second mirror ($l = L$) a certain fraction of the radiation is transmitted and the remaining fraction R is reflected. (For plane waves vertically incident on the GaAs/air interface the value of R is about 0.32.) For the oscillation to be self-sustaining the reflected wave must in turn have at least the intensity I_0, the oscillation condition then being

$$R \exp(\Gamma G - \alpha)L = 1. \qquad (2a)$$

This can also be written as:

$$\Gamma G = \alpha + \frac{1}{L} \ln \frac{1}{R}. \qquad (2b)$$

The value G_{th} of G that meets these conditions corresponds to the threshold current. As can be seen, G_{th} decreases with the length of the laser. In itself this is

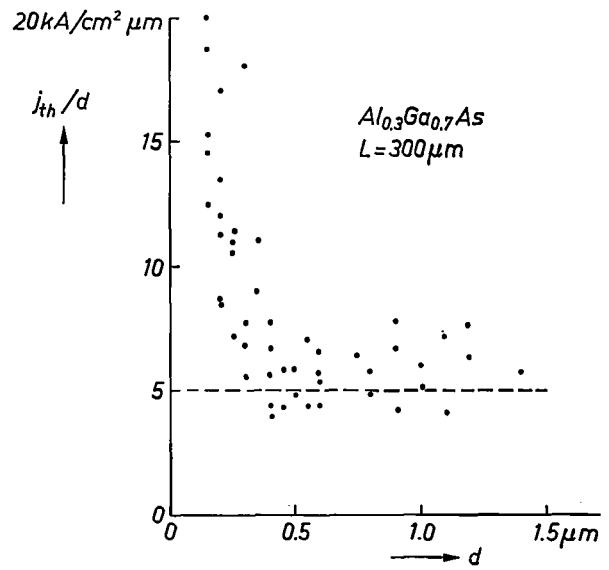


Fig. 5. Ratio of the density j_{th} of the threshold current to the thickness d of the active layer of a number of experimental lasers (with $x = 0.3$ and a length of 300 μm) as a function of d. Below about 0.3 μm this ratio increases sharply; above it j_{th}/d is approximately constant. The dashed line gives the approximate theoretical value (5 kA/cm² μm).

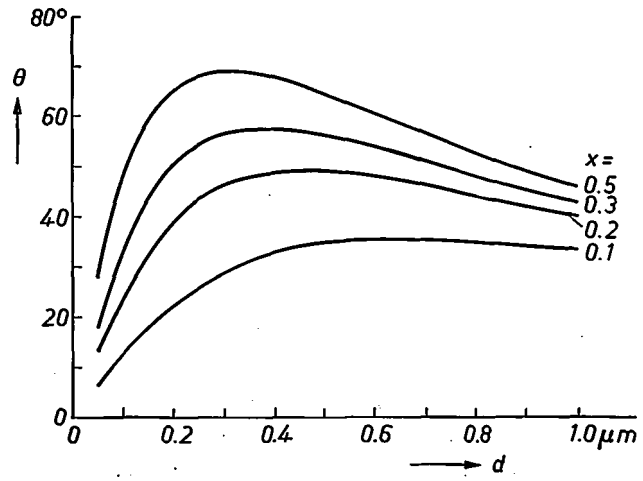


Fig. 6. The 50% aperture θ of the exit laser beam (in the direction perpendicular to the active layer) as a function of d for four values of the relative aluminium content x (calculated).

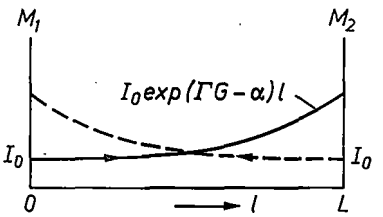


Fig. 7. Illustrating the condition under which a laser can operate (oscillation condition). $M_{1,2}$ mirrors. L length of the laser. The solid curve shows how a wave of intensity I_0 beginning at M_1 increases exponentially in intensity with the distance l travelled to the other mirror. The oscillation is sustained only if the wave reflected from M_2 is at least of intensity I_0. G gain and α attenuation in the active layer per unit pathlength. Γ is the fraction of the total power in the wave transmitted through the active layer.

[3] G. A. Acket, W. Nijman and H. 't Lam, J. appl. Phys. 45, 3033, 1974.
[4] The idea that steps in the refractive index can be used for reducing the threshold current was put forward by G. Diemer and B. Bölger of Philips Research Laboratories (Physica 29, 600, 1963).
[5] W. W. Anderson, IEEE J. QE-1, 228, 1965.
[6] M. J. Adams and M. Cross, Electronics Letters 7, 569, 1971.
[7] P. J. de Waard, Electronics Letters 11, 11, 1975.

by no means a surprising result, but the equation makes it possible to determine the value of L above which no significant improvement can be achieved; this value is about 300 μm.

If the current through the laser exceeds the threshold value, nearly all the additionally injected charge carriers recombine by stimulated emission. The efficiency η_{int} of this process, the 'internal' efficiency, is between 60 and 100%. The stimulated emission is divided between the internal losses αL and the mirror losses $\ln(R^{-1})$, so that the external differential quantum efficiency η_{ext} per mirror is given by:

$$\eta_{ext} = \frac{\eta_{int}}{2} \frac{\ln(R^{-1})}{\alpha L + \ln(R^{-1})}. \quad (3)$$

Fig. 8 shows how the intensity of the exit light beam varies with the current through the diode. Above the threshold current the relation is practically linear, with a slope coefficient $E_g\eta_{ext}$.

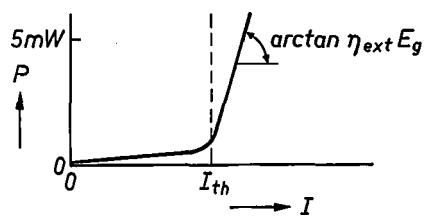


Fig. 8. Light-current characteristic of a laser. If the current is higher than the threshold value I_{th} for laser action, the emitted power P increases much more rapidly with I than below this value. E_g band gap. η_{ext} differential quantum efficiency per mirror (see eq. 3).

Apart from the power requirement (2) it is of course also necessary to meet the optical requirement that L must be equal to a whole number (m) of half-wavelengths (Fabry-Pérot):

$$m\lambda/2n = L. \quad (4)$$

It follows from the values of L, n and λ given that this number is about 2500. Since the wavelength region in which laser action is possible has a certain width, there are various wavelengths for which equation (4) is satisfied with such a long laser, but one of these is dominant if the current density is not unduly high. These wavelengths lie about 0.3 nm apart.

Liquid-phase epitaxy of double-heterojunction lasers

The lasers we use are made by means of liquid-phase epitaxy (LPE). In the LPE process a thin single-crystal layer, often only a few μm thick, is deposited from the liquid phase on to a carefully prepared single-crystal substrate. The crystal structure of the substrate determines that of the epitaxial layer. A distinction is made between *homo*-epitaxy, where the layer has the same chemical composition as the substrate (except for the nature and concentration of the doping elements), and *hetero*-epitaxy, where there is a distinct difference in the compositions of the layer and the substrate. A device like that of fig. 2 is therefore made by hetero-epitaxy.

To illustrate the principle of LPE and our own working procedure we shall take the simplest case, i.e. epitaxial growth of GaAs from a Ga-rich liquid phase on a GaAs substrate. *Fig. 9* shows a simplified phase diagram of the binary III-V system Ga-As. The metallic Ga melts at about 29 °C and As at about 818 °C. The two elements form a compound, GaAs, which has a very narrow existence region. The deviation from the stoichiometric composition is very small: $\delta y < 10^{-4}$. The LPE now takes place from a liquid which is characterized in fig. 9 by the point y_1, T_1. A liquid of the composition y_1 can be considered as a solution of As in Ga. At the temperature T_1 the solution is in equilibrium with solid GaAs. If the temperature is reduced to T_2, the solution becomes supersaturated, because the equilibrium concentration associated with T_2 is smaller than y_1. What then happens depends on the magnitude of $T_2 - T_1$. If this difference is 0.25 °C or less, nothing will happen unless a slice of GaAs is immersed in the liquid. A layer of GaAs then settles

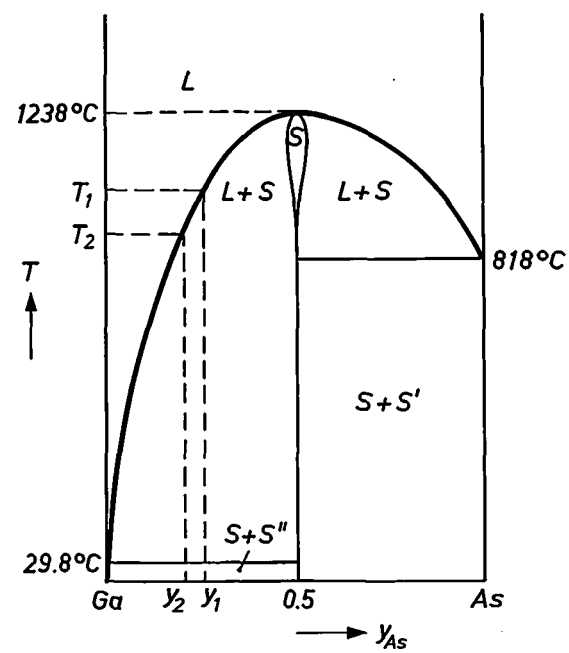


Fig. 9. Phase diagram of the gallium-arsenic system (schematic). L liquid. S solid. The compound GaAs has only a very narrow existence region ($\delta y < 10^{-4}$). The epitaxial deposition, starting from the temperature T_1 at which the substrate of GaAs and the liquid (with composition y_1) are in equilibrium with one another, takes place by reducing the temperature by a suitably small amount to T_2. The liquid then becomes supersaturated ($y > y_2$) and a layer of GaAs settles on the substrate.

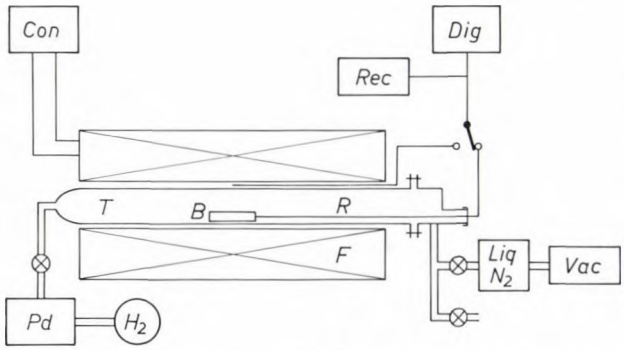


Fig. 10. Arrangement for liquid-phase epitaxy (LPE) for making GaAs lasers with a double heterojunction. Hydrogen delivered by cylinder H_2 and the purification system *Pd* (with palladium filters) flows through the tube *T*. *F* three-zone furnace, which can be set up to give the desired temperature profile. The control system *Con* keeps the temperature constant to 0.1 °C and also provides for a very slow linear decrease of the temperature with time. The epitaxial process takes place in the boat *B*, operated by the rod *R* (see fig. 11). At various places the temperature can be measured with thermoelements, which are connected to the recorder *Rec* and the digital voltmeter *Dig*. To the right of *T*, connected in parallel, are an outlet for the hydrogen gas (below) and a rotary vacuum pump *Vac* with liquid-nitrogen cooler (*Liq* N_2).

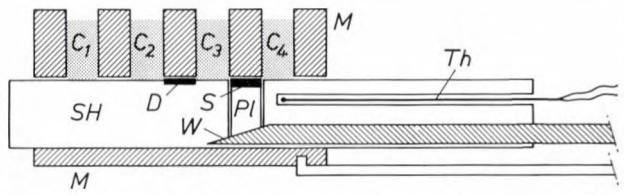


Fig. 11. Multiple-bin boat for multiple-layer epitaxy (*B* in fig. 10). *M* graphite block with four cavities C_{1-4} containing the liquids to be used. *SH* substrate holder, a second graphite block that divides off the spaces C_{1-4} from the others. The substrate *S* can be slid in turn under C_1 to C_4. Accurate external control of the substrate height is carried out with a wedge *W* with a plunger *Pl*. *D* disc of GaAs (source-seed substrate) which is brought into contact with the liquids before the actual substrate, to eliminate the effect of a possible mismatch between T_1 and y_1 (see fig. 9). *Th* thermoelement.

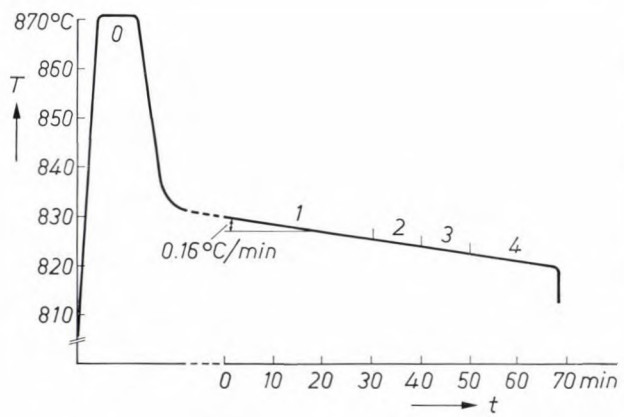


Fig. 12. Temperature-time diagram for multilayer epitaxy with the arrangement in fig. 10. The four layers are successively deposited in the periods *1* to *4*. Period *4* is ended by pulling out the substrate from under the last solution and switching off the furnace; note the sharp drop in temperature. Before the depositions take place the whole assembly is kept for some time at a somewhat higher temperature to homogenize the solutions (period *0*).

on this slice alone; this is the epitaxial deposition referred to. At greater $T_2 - T_1$ deposition also takes place on the walls of the vessel or 'bin' that contains the liquid, and at still greater $T_2 - T_1$, e.g. 20 °C, nucleation takes place throughout the liquid. The last two effects are of course unwanted.

Fig. 10 gives a diagram of the arrangement for the multilayer epitaxy of GaAs/AlGaAs. A three-zone furnace *F* contains a reactor tube *T* through which purified hydrogen gas flows. Inside the reactor is the boat *B* in which the epitaxial deposition takes place. The temperatures can be very accurately adjusted to the desired temperature profiles by means of the furnace, and a control unit is used to reduce the temperature slowly and linearly with time. The hydrogen gas contains less than 1 ppm of water vapour; a water content as low as this is necessary because aluminium has a great affinity to oxygen and because the oxide is stable to far beyond the temperature at which the deposition takes place. The presence of aluminium oxide can give rise to crystal defects and to nucleation errors.

The boat is not a simple container but a complicated structure, whose cross-section is shown schematically in *fig. 11* [8]. In this 'multiple-bin boat', as it is called, the layers from which the laser is built up can be directly applied one after the other. There are two main parts. The first is a block of graphite *M* containing a number of holes C_i (four in this case) which act as vessels that hold the various solutions required for the production. The base of these vessels is formed by a second block *SH*, which can be slid under the first. This block serves as substrate holder, enabling the substrate to be slid successively under C_1, C_2, etc. In most of our boats the graphite blocks are coated with a thin layer of pyrolytic graphite [9].

The time that the substrate remains under a particular solution and the cooling rate are the main factors that determine the thickness of the resultant layer. *Fig. 12* shows a temperature-time diagram of a complete growth process.

Multiple-bin boats are used for many kinds of experiment in many countries; they differ in their details, however, depending on the type of work. To make a long-life laser a very high crystal perfection is needed — there must be no dislocations. It must also be possible to control the composition of the various layers very accurately, and also the thickness of the layers, since this determines a number of essential quantities such as the threshold current density and the

[8] T. G. J. van Oirschot and W. Nijman, J. Crystal Growth **20**, 301, 1973.

[9] See for example W. F. Knippenberg, B. Lersmacher, H. Lydtin and A. W. Moore, Philips tech. Rev. **28**, 231, 1967.

thermal resistance. As explained above, the transition from the P to the N region must be very abrupt (within 50 nm), which means that when the substrate is slid along to the next compartment C_i it must carry virtually no liquid with it. The interfaces between the layers must in addition be optically flat (no ridges). The surface of the substrate must therefore closely coincide with a low-index crystal plane. In our substrates this surface coincides within 0.4° with a {001} plane. The dislocation density in our substrate is lower than 5000 cm^{-2}.

The ability to make long-life lasers is due not only to the LPE equipment described in figs. 10 and 11 but also to the good match between the crystal lattices of GaAs and AlGaAs. *Fig. 13* shows the temperature dependence of the lattice constants of GaAs, AlAs and $Al_{0.3}Ga_{0.7}As$. It can be seen that the values are nearly identical at the deposition temperature, an almost ideal situation. At room temperature the greatest relative

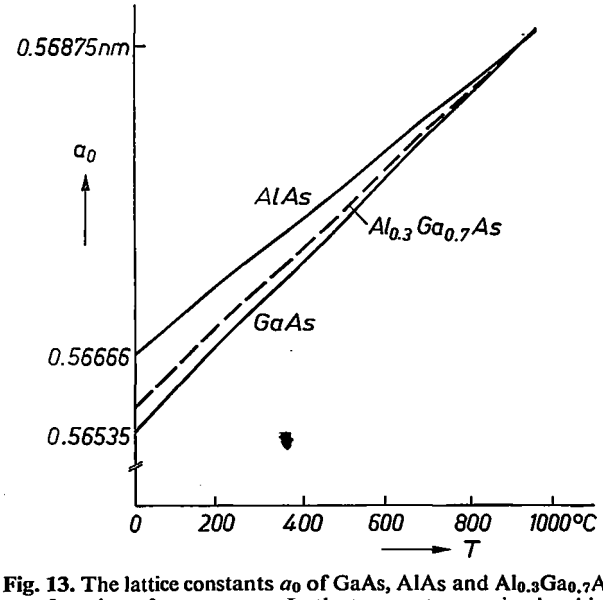


Fig. 13. The lattice constants a_0 of GaAs, AlAs and $Al_{0.3}Ga_{0.7}As$ as a function of temperature. In the temperature region in which the epitaxial deposition takes place they are almost identical, at room temperature the relative temperature difference is no more than 0.14%. It is partly because of these very favourable circumstances that long-life continuous-wave GaAs lasers can be made.

difference is about 1.4×10^{-3}. This difference is so small that it can be taken up by elastic deformation in layers of the thickness found here, without the appearance of any dislocations. The elastic stress thus produced is about 3000 N/cm². Care needs to be taken, however, that the stress caused by mounting does not make the total stress so high that there is plastic deformation.

The mismatch between the lattice constants of other (ternary or quaternary) III-V compounds that might be suitable as laser materials — e.g. $Ga_{1-x}In_xAs_{1-y}P_y$ or $Ga_{1-x}Al_xAs_{1-y}Sb_y$ — is a serious obstacle to the production of long-life continuous-wave lasers with such materials. They can be used for making lasers with wavelengths of 0.7-1.7 μm. Those with a wavelength of about 1.1 μm would be very attractive for optical communication [10].

The doping elements used for making N- and P-type material are usually tin (for N) and germanium (for P). These elements are not highly volatile at 800 °C and diffuse extremely slowly within the solid. They are therefore particularly suitable for making the very steep P-N junctions required. Another practical advantage is that in the Ga/GaAs system the segregation coefficient for very many elements is of the order of only 10^{-3}, so that minor contamination of the solution does not have any noticeably adverse effects on the purity of the deposited layer.

Apart from random deviations, the initial temperature of the furnace will not in practice always correspond exactly to the mass ratio of the chemicals at the beginning, either because of a deviation in the thermoelectric potential of the thermoelements that control the furnace temperature, for example, or because of small weighing errors. Since the deposited layer must be no more than a few μm thick, a temperature deviation of a few tenths of a degree in a fixed growth time is enough to cause a relatively marked difference in thickness.

To get around this difficulty the liquid, prior to the deposition on the laser substrate, is left in contact for some time with another substrate of the same composition [11]. If there is any discrepancy between composition and temperature, some material will settle or go into solution, depending on the sign of the deviation, until a perfect match is achieved. The laser substrate is thus brought into contact with the liquid. With this 'source-seed technique', as it is called, the spread in the thickness of the active layer can be limited to 0.1 μm. *Fig. 14* shows a photomicrograph of a cleavage plane, perpendicular to the four layers.

Peltier LPE

Besides the method of liquid-phase epitaxy described above, it is interesting to consider a deposition method in which the furnace is kept at a constant temperature — and hence also the liquid and the substrate. In this method a direct current is passed through the liquid-substrate interface, where it causes cooling due to the Peltier effect. The cooling brings about the deposition on the substrate. This method has previously been tried with InSb, GaAs and AlGaAs [12–15]. The experiments with AlGaAs showed that very homogeneous layers of $Al_xGa_{1-x}As$ can be made in this way (the

spread in x was less than 0.005 over a thickness of 60 μm [14]) and also that the growth rate is proportional to the current [13–15].

Since the growth rate is determined entirely by the current, the thickness of each of the deposited layers is easily controlled. In the epitaxial deposition of AlGaAs we were also able to work with very low growth rates [14] since the deposition can take place at a lower temperature (600 °C) than is usual in LPE. This is an additional advantage in that it can lead to lower point-defect concentrations.

Fig. 15 shows a schematic cross-section of the boat for Peltier LPE, made from graphite and boron nitride. Four liquids are used, and here again a source-seed wafer precedes the deposition. A different current is used for each of the four layers, so that each is formed at the optimum growth rate. A characteristic example of a deposition at 800 °C is given in *Table I*.

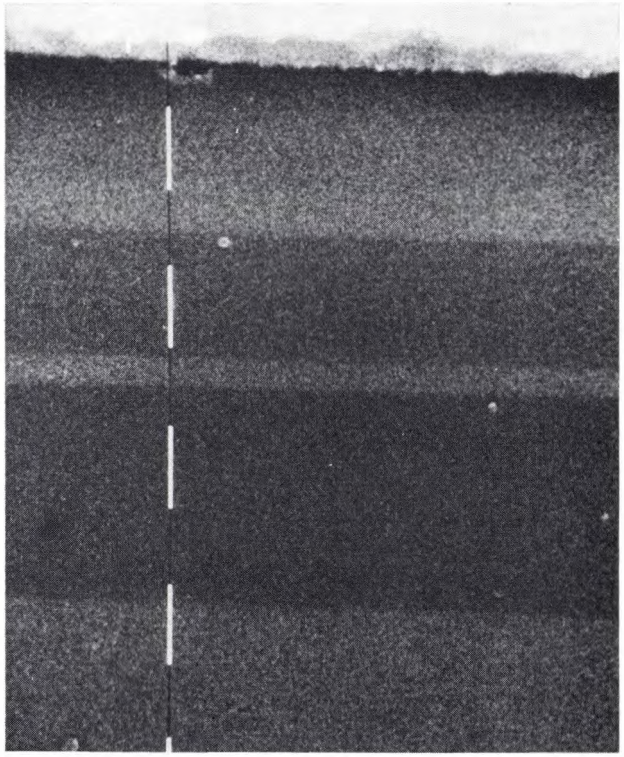


Fig. 14. Photomicrograph taken with a scanning electron microscope (PSEM 500) of a multilayer system as in fig. 2, made by liquid-phase epitaxy. The vertical white strips are each 1 μm long.

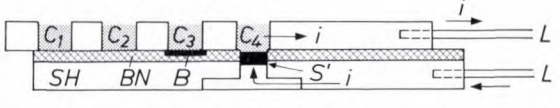


Fig. 15. LPE boat in which the temperature reduction required for the deposition is bought about by passing a direct current i through the liquid-substrate interface (Peltier LPE). One of the differences compared with the boat in fig. 11 is that the block *SH* is coated with an insulating layer *BN* of boron nitride. *L* stainless-steel leads for current input and output. The other symbols have the same significance as in fig. 11.

Table I. Some characteristic quantities in the application of four epitaxial layers (see fig. 2) by Peltier LPE at 800 °C.

Layer	Thickness (μm)	Current (A)	Growth time (min)	Growth rate (nm/min)	Luminescence peak (nm)
1	7.5	8.0	25	300	669
2	0.24	0.25	14	17	880
3	1.4	8.0	30	47	684
4	2.6	3.0	10	260	872

Test measurements

As has been explained above (see equations 1), some of the material parameters of multilayer systems of $Al_xGa_{1-x}As$-GaAs on which laser action depends are the thickness of the (future) active layer, the aluminium content of both the active layer and the two adjoining layers, and the lifetime τ for spontaneous recombination of the minority charge carriers in the active layer.

The thickness of the active layer can be measured with a scanning electron microscope. The aluminium content of the various layers is determined by means of photoluminescence. Since the electrons and holes released by the incident short-wavelength radiation occupy energy levels close to the lower edge of the conduction band and the top of the valence band respectively, the quantum energy of the luminescence radiation is approximately equal to the band gap. This in its turn is a measure of the aluminium content.

To investigate our multilayer systems by this method our wafers are cleaved along a plane perpendicular to the layers and illuminated through a microscope objective, which focuses the beam of a helium-neon laser to an extremely small light spot. A complete luminescence spectrum is then recorded at various points of the cleavage plane. From the positions of the light spot and the wavelength of the peaks in the spectra the composition of the layers can be derived. *Fig. 16* shows a family of luminescence spectra obtained in this way from a wafer with an active layer consisting of $Al_{0.1}Ga_{0.9}As$.

The lifetime τ is also determined by means of photoluminescence. In this case the laser is modulated at a high frequency. The corresponding intensity variation of the luminescence radiation, which depends on τ, is analysed with a very fast photomultiplier [3].

[10] D. N. Payne and W. A. Gambling, Electronics Letters **11**, 176, 1975.
[11] L. R. Dawson, J. Crystal Growth **27**, 86, 1974.
[12] M. Kumagawa, A. F. Witt, M. Lichtensteiger and H. C. Gatos, J. Electrochem. Soc. **120**, 583, 1973.
[13] J. J. Daniele and C. Michel, in: Gallium arsenide and related compounds, Proc. Conf. Deauville 1974 (Inst. Phys. Conf. Ser. No. 24), p. 155.
[14] J. J. Daniele, Appl. Phys. Letters **27**, 373, 1975.
[15] D. J. Lawrence and L. F. Eastman, J. Crystal Growth **30**, 267, 1975.

Fabrication of laser diodes

Lasers for optical communication must not be larger than about 300 μm (length) by about 5 μm (width) in the plane of the *P-N* junction. Since pieces of material as small as this would be extremely difficult to handle, we work with 'dices' measuring 300 × 300 μm. with metal, doping of 10^{18} at/cm^3 is sufficient. On the other side, where a stripe of only 5 μm is conductive, stricter measures need to be taken. Not only is the Ge doping twice as strong here, but the acceptor concentration is also increased (up to 10^{19} cm^{-3}) by diffusing zinc atoms into the layer at 600 °C.

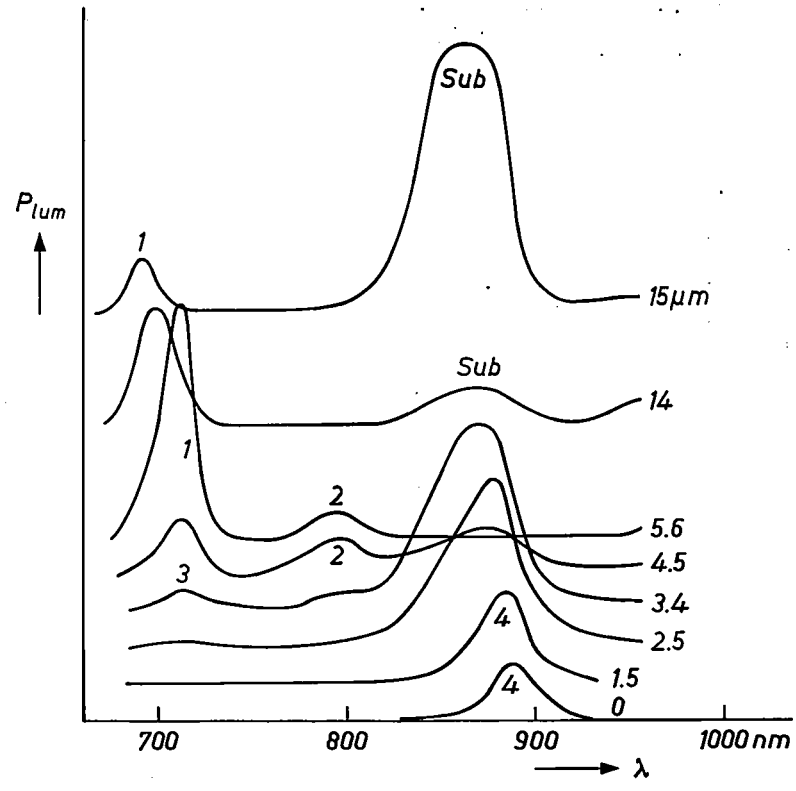


Fig. 16. The composition of the epitaxial layers is determined by measuring the photoluminescence spectrum upon excitation with a HeNe laser beam sharply focused on a number of positions on a cleavage plane. The positions, measured from the surface of the upper layer, are indicated beside the spectra. The luminescence peaks are indicated by the symbol of the layer to which they correspond (see fig. 2). The band gap of the appropriate layer is derived from the positions of the peak; this is a measure of the aluminium content.

Two strips of nearly 150 μm as seen in the transverse direction are made non-conducting by bombarding them with protons of the appropriate energy [16]; a 'stripe' of 5 μm between them is left intact by protecting it with a gold layer. Care is of course taken to ensure that this stripe is perpendicular to the cleavage planes that must act as reflectors.

The electrical contacts have to provide not only for the passage of the electric current but also for heat conduction, i.e. for the removal of most of the energy supplied. To obtain a virtually resistance-free transition from the metal electrodes to the GaAs to prevent heat generation in the contacts, the top or 'capping' layer and the substrate are both heavily doped. Since the whole substrate is conductive and completely covered

To ensure the maximum heat conduction the contacts are made of gold and the laser is soldered to a copper block with the epitaxial layers against the block; in this way the block is closest to the place where the heat is generated, and the resistance between the heat source and the environment is kept sufficiently low to make continuous laser operation possible. An additional advantage of gold as a contact material is that it can be made fairly soft by heat treatment, so that the difference in the coefficients of expansion of contact material and GaAs does not cause unduly high mechanical stresses in the laser. Indium, another soft material, is used for soldering the laser to the copper block. *Fig. 17* shows a scanning-electron photomicrograph of a laser mounted in the manner described.

Life of a continuous-wave laser

In continuous-wave operation a GaAs laser gradually suffers degradation, i.e. some of its properties deteriorate. These changes are not reversible, and eventually the laser becomes unusable. The most striking change is an increase in the threshold current. *Fig. 18* shows a family of light/current characteristics of a c.w. laser, measured at intervals of 100 to 200 hours. After about 1700 hours the laser was no longer capable of c.w. operation. It can also be seen from the figure that not only does the threshold current increase but the curves gradually become less steep, that is to say η_{ext} decreases. Both effects may be the result of increasing attenuation α; see (2b) and (3) — G varies linearly with j_{th}.

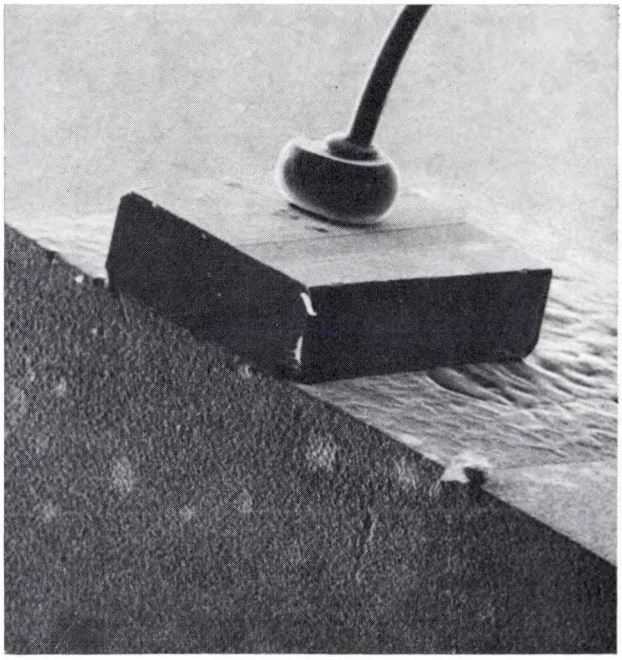


Fig. 17. Experimental GaAs laser — substrate with four epitaxial layers — soldered to a copper block for heat dissipation. The four layers are located on the side of the block because the heat generation is closest to the block in this position. A strip has been removed from the substrate contact for observations (see figs. 19 and 20).

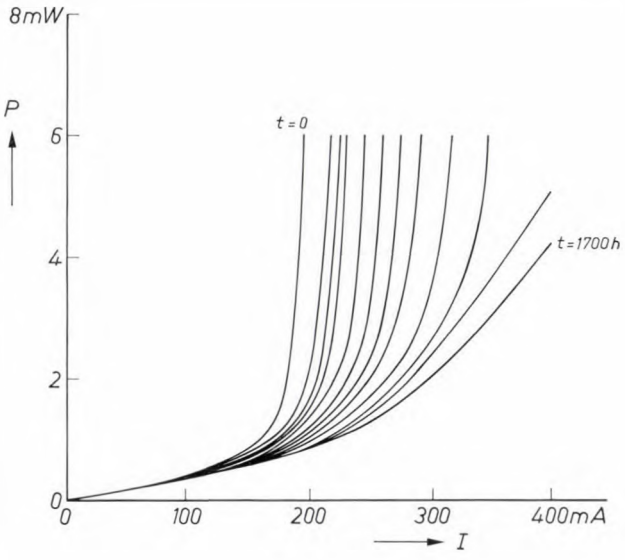


Fig. 18. Family of light/current characteristics, measured during a life test with one of the experimental GaAs lasers. Note that the threshold current I_{th} for laser action (see fig. 8) gradually increases and that the slope of the curves gradually decreases, i.e. the external efficiency deteriorates.

Until the present, investigations into the causes of the degradation have been mainly phenomenological. Although they have provided a good picture of what takes place in the ageing laser, they have not yet given a complete description of all the effects involved. Nevertheless the useful life of the laser has in the meantime been increased to about 5000 hours.

Fig. 19 and *fig. 20* show infrared photomicrographs of the interior of lasers that have very probably become degraded by the growth of dislocation networks [17]. In fig. 19 a number of parallel dark lines can be seen to the right of the centre at an angle of about 45° to the longitudinal axis. They lie in one of the crystal planes, which is often the case with dark regions of this type. This is one of the reasons for thinking that they may be caused by dislocation networks. Thermal measurements have shown that the temperature in the dark regions is some tens of degrees higher than elsewhere [18]. This temperature difference and the lower emission of radiation can both be explained from the presence of dislocations. In the first place there is a greater probability of radiationless recombination at dislocations, causing less radiation to be emitted there than elsewhere. Dislocations also absorb more radiation, thereby adding to the heat generation.

The dislocation networks in the laser of fig. 19 are presumably due to mechanical damage during mounting. The dark regions in the laser in fig. 20, on the other hand, are caused by dislocation networks which probably had their origin in the epitaxial layers themselves. In both cases the network gradually spreads out under the influence of mechanical stresses caused by temperature differences.

[16] L. A. D'Asaro, J. Luminescence **7**, 310, 1973.
[17] This explanation was first put forward by P. Petroff and R. L. Hartman, Appl. Phys. Letters **23**, 469, 1973.
[18] T. Kobayashi, T. Kawakami and Y. Furukawa, Jap. J. appl. Phys. **14**, 508, 1975.

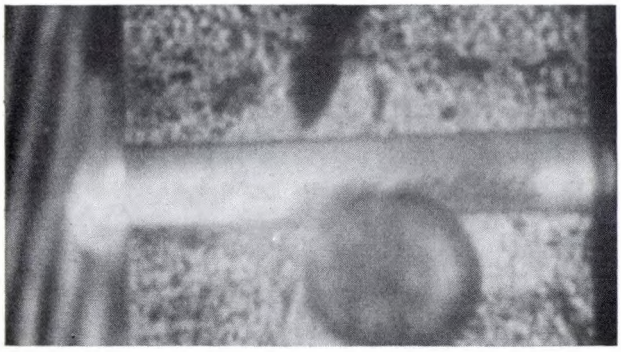


Fig. 19. Photomicrograph, taken with infrared microscope, of the spontaneous luminescence in the active layer of an internally damaged laser. To the right of centre a series of parallel dark stripes can be seen; these are most probably due to the presence of the same number of dislocation networks on crystal planes.

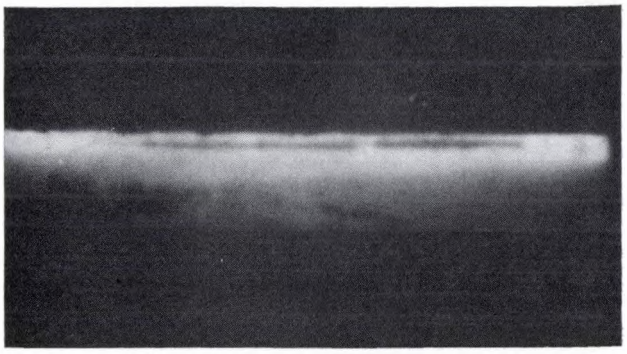


Fig. 20. As fig. 19. The dark regions, lying in the longitudinal direction of the laser, are presumably due in this case to dislocation networks originating in one of the epitaxial layers.

Apart from the internal cause shown in figs. 19 and 20, damage to the mirrors can also cause the deterioration of a laser. In some lasers the quality of the mirrors gradually deteriorates, for reasons that are not yet fully known. It is of course obvious that care should be taken to keep the mirrors out of contact with substances that attack the GaAs, such as water vapour.

If the drive current is far above the threshold value the mirrors may be destroyed within a few seconds, either by breakdown due to excessive (optical) field-strength (> 20 kV/cm) or by an excessively high local temperature rise [19]; temperatures above 1000 °C have been observed. Examples of such damage can be seen in *fig. 21*.

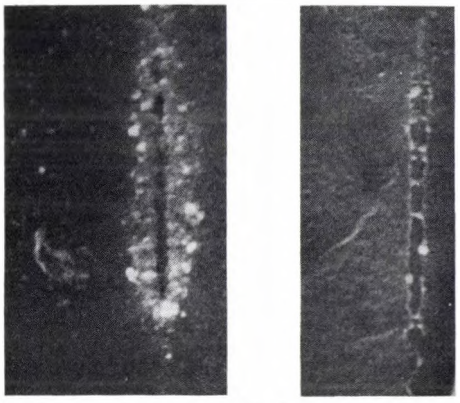
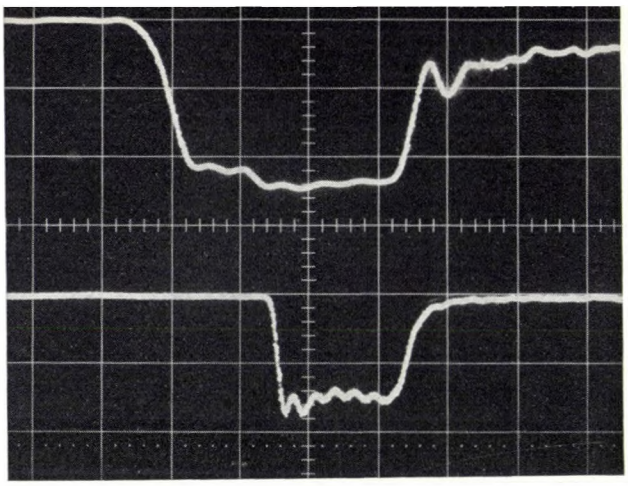


Fig. 21. *Left:* photomicrograph of a laser mirror which was very quickly destroyed by an excessively high drive current. Note the row of craters (width about 0.5 μm), presumably due to electrical breakdown. *Right:* Damage to a mirror whose quality has very gradually decreased in the course of time. The cause of this kind of degradation is not yet fully understood.

[19] B. W. Hakki and F. R. Nash, J. appl. Phys. **45**, 3907, 1974.

Summary. A light source that seems very suitable for optical communication with glass fibres, notably because of its smaller dimensions and ease of modulation, is the solid-state laser. A GaAs/$Al_xGa_{1-x}As$ (diode) laser with a double heterojunction has been developed that requires such a low drive current (100-200 mA) that it is capable of continuous-wave operation at room temperature. Four layers are deposited epitaxially on an *N*-type GaAs substrate: an $Al_xGa_{1-x}As$ layer (*N*), a GaAs layer (*P*), an $Al_xGa_{1-x}As$ layer (*P*) and another GaAs layer (*P*). The laser action takes place in the first *P* layer (thickness about 0.3 μm), in which a strip about 5 μm wide is active; the rest of the layer is made inactive by bombarding the layers above it with protons. Cleavage planes of the crystal act as laser mirrors. The radiation is confined to the direction perpendicular to the layers because AlGaAs has a different refractive index ($\Delta n \approx 0.7x$). The layers are made by liquid-phase epitaxy with a multiple-bin type of boat either by gradually cooling the boat or by cooling the solid-liquid interface by passing a direct current through it (Peltier epitaxy). Blocks measuring 300×300 μm are soldered with indium to a copper block for heat dissipation. After a period of continuous operation the laser gradually deteriorates in quality owing to the formation of dislocation networks or to damage of the mirrors. A useful life of 5000 hours for these lasers has already been achieved.



uctor lasers

The figure on page 201 in this issue of Philips tech. Rev., Volume **36**, has unfortunately been printed upside down. Please affix this sticker in the correct position.

The most suitable light source for the large bandwidth potential of optical communication by means of glass fibres is probably the double-heterojunction $GaAs/Al_xGa_{1-x}As$ laser, described elsewhere in this issue [1]. This semiconductor laser is of interest as a signal source because it is small and its radiation can be modulated by varying the injection current so that the information to be conveyed can be incorporated directly into a light signal. With pulse modulation this method can be used up to about 2 Gbit/s. To ensure interference-free signal transfer over a considerable distance, the energy of the light pulses should of course be made as large as possible. The amplitude of a pulse at the end of the fibre will in any case have to exceed a certain minimum value to ensure that the detector does not mistake a '1' for a '0'.

The more energy there is in the laser pulse and the more efficient the coupling to the fibre, the greater the spacing permissible between repeaters, provided that pulse dispersion is not the limiting factor (alternatively: the easier it will be to maintain the amplitude above the minimum value); the coupling efficiency is particularly important because there are limits to the luminous flux that a laser can deliver. The energy in a pulse does not only depend on the duration and amplitude of the electrical drive pulse but is also determined by the temperature of the laser. In this article we shall discuss a method for reducing this unwanted effect and also our solution to the problem of coupling glass fibres to lasers.

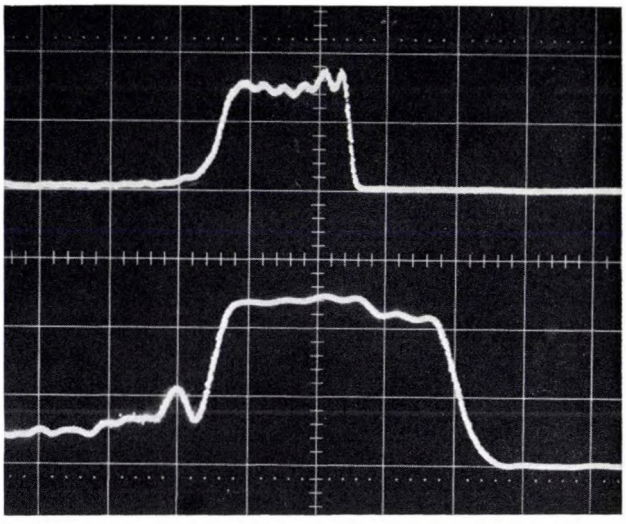


Fig. 1. The initial delay. *Above:* pulse-modulated injection current in a GaAs laser. *Below:* the laser pulse resulting from the current pulse; the laser pulse is shorter, chiefly because it begins later. One scale division is equivalent to 1 ns. The initial delay can be overcome by using a 'bias current' which is equal to the threshold current with superimposed modulation pulses.

Modulation

Laser light can in principle be modulated by switching the injection current on and off. Generally, however, a constant current which is approximately equal to the threshold current is passed through the laser and the modulation pulses are superimposed on this 'bias current'. This is done because the start of the laser pulse is delayed with respect to the current pulse; the delay can be as much as several nanoseconds if the bias current is small or zero (*fig. 1*). The delay arises because the occupation of the inversion layer cannot rise instantaneously to the threshold value required for laser operation [1]. Moreover, spontaneous recombinations occur all the time in the active zone of the laser, and these have the effect of further reducing the rate of occupation. Modulation rates in the region of 2 Gbit/s would not be attainable without bias current. The initial delay is undesirable even at lower modulation rates since it reduces the energy in the laser pulse. The initial delay makes the laser pulse shorter than the current pulse, since both end at practically the same instant. The less the bias current differs from the threshold current the shorter the delay. *Fig. 2* shows that the light signal from a laser made in our laboratories, with the bias current equal to the threshold current, follows the electrical modulation signal very closely even at a modulation rate of 500 Mbit/s.

Ir G. D. Khoe and L. J. Meuleman are with Philips Research Laboratories, Eindhoven.

[1] The operation of a solid-state laser is discussed in the article by G. A. Acket *et al.*, this issue, p. 190.

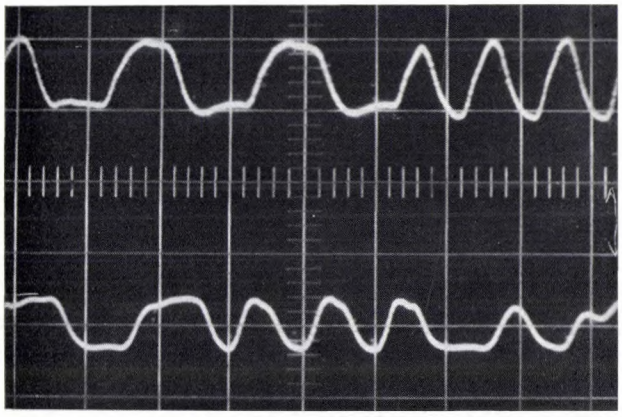


Fig. 2. An example of a modulated light signal (top) from a GaAs laser when the bias current is equal to the threshold current. The digital signal shown conveys information at a rate of 500 Mbit/s; the signal was measured after transmission along an optical fibre with a length of 0.5 metre. The lower oscillogram shows the current. One scale division here is equivalent to 4 ns.

An expression for the initial delay Δt of the laser pulse can be derived from the equilibrium equation for injected minority carriers [2]. We then find:

$$\Delta t = \tau \left(\ln \frac{I_m - I_b}{I_m - I_{th}} \right),$$

where I_{th} is the threshold current, I_b the bias current and I_m the value of the injection current after a step in the bias current; τ (about 5 ns) is the average lifetime of injected charge carriers if they only recombine spontaneously. It follows from the relation given, which applies for $I_m > I_{th}$, that the initial delay disappears when the bias current is made equal to the threshold current.

The endeavour to produce powerful, constant light pulses is complicated by the problem of the temperature dependence of the threshold current: a higher temperature necessitates a higher threshold current. For constant bias current and constant amplitude of the electrical pulse, the energy of the light pulse therefore decreases with increasing temperature. This effect is to some extent enhanced because the initial delay also increases. We have found a way out of this difficulty by deriving a control signal from the laser light produced and using it to control the bias current. A first experiment has shown that by using this feedback principle the drop in energy of the light pulses can be reduced to below $\frac{1}{4}\%$ per °C rise in temperature.

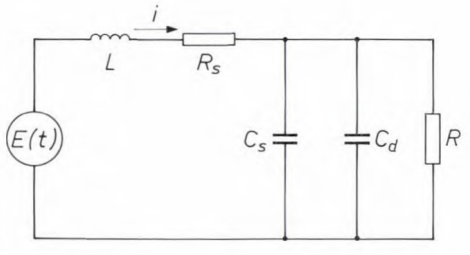


Fig. 3. A simplified equivalent circuit for a GaAs laser with its power supply. $E(t)$ the time-dependent voltage delivered by the power supply. R_s sum of the internal impedance of the power supply and the series resistance of the laser. L series inductance originating in the connections between the power supply and the laser. C_s space-charge capacitance of the depletion layer: this depends on the voltage across the layer. C_d diffusion capacitance of the depletion layer; this depends on the injection current i. R diode resistance; this depends on the injection current.

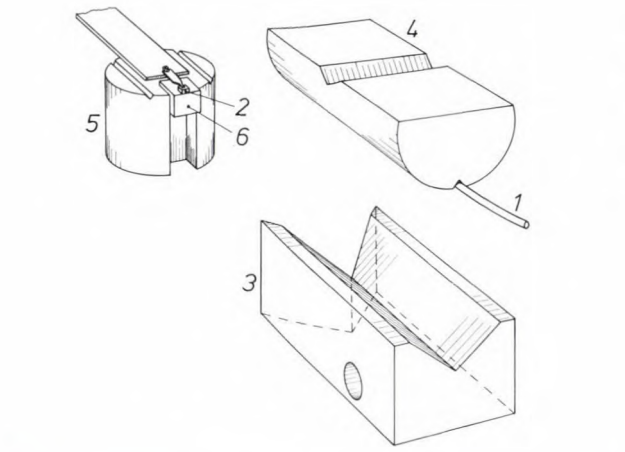
In designing a power supply for the laser we started by analysing its behaviour as an element in an electrical network, using the equivalent circuit shown in *fig. 3*. The various impedances are 'dynamic' quantities in the sense that they depend on the currents and voltages in the laser. The choice of the internal impedance of the power supply was based on the results of this analysis. Careful assembly, with special attention to minimizing unwanted series inductance, resulted in a circuit that will provide both the bias current and a modulation signal up to a maximum rate of 500 Mbit/s.

Coupling

Fig. 4 shows a universal device that we have designed for coupling lasers to fibres. It is suitable for single-mode and multimode fibres [3]. The fibre and laser are placed on cylindrical carriers that are positioned relatively to an accuracy of about 0.2 μm with the aid of a micromanipulator before adhesive is applied. This great accuracy is necessary because the core of a single-mode fibre is only about 3 μm across. The positioning requirements for coupling to a multimode fibre — core diameter of 30-80 μm — are far less exacting.

With a multimode fibre, in which the laser light can excite a large number of electromagnetic modes, direct coupling will give very inefficient energy transfer. This is because the exit beam diverges markedly in the plane perpendicular to the active layer in the laser, so that much of the light arrives in the fibre at an angle that makes total internal reflection impossible. In fibres with a graded index, i.e. fibres in which the refractive index decreases gradually with the radial coordinate, a considerable amount of light would be lost because the acceptance angle of the fibre is not uniform over the entire entrance plane, but decreases with the radius of the point where the light enters. We avoid these difficulties by inserting a suitable lens in front of the entrance plane of the fibre. The lens converges the laser light to an almost parallel beam, mainly incident within the acceptance angle of the fibre (*fig. 5*). Since the light is parallel, only a few modes are excited in the fibre and the dispersion is reduced [4].

For wideband communication over large distances it is preferable to use single-mode fibres. The laser should then operate in its lowest-order transverse mode. Theoretical considerations [5] show that the highest



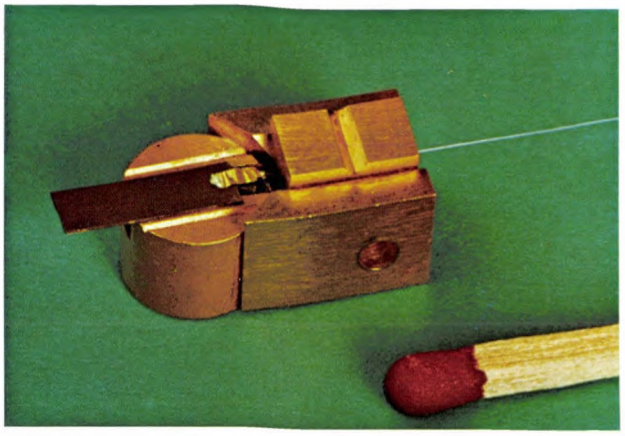


Fig. 4. A universal device for coupling a single-mode or multimode fibre (*1*) to a GaAs laser (*2*). The coupler (*3*) contains grooves in which two carriers (*4* and *5*) can be fastened securely once the active layer and the fibre core have been accurately aligned with each other, by sliding and rotating the carriers. A small rotation of the fibre carrier, for instance, will move the fibre practically horizontally because of the eccentric position of the fibre in the carrier. A small block (*6*) under the laser acts as a positive electrode and a heat sink. The strip-shaped electrode of the laser is on the underside of the block, which aids heat transfer to the block. The (outer) diameter of the fibre is 100 μm. The active zone in the laser is about 5 μm wide and 0.5 μm high. The dimensions of the actual coupling device, as the colour photograph shows, are small. A slight modification to the coupling block allows two fibres to be coupled to a laser (see also fig. 6).

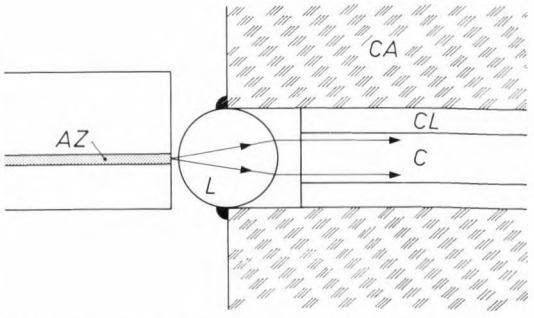


Fig. 5. The light beam leaving a GaAs laser has a marked divergence perpendicular to the active zone (*AZ*). This causes excessive loss of light when the laser is coupled to a multimode fibre. A simple solution is to place a spherical lens (*L*) in front of the entrance plane in the fibre. The lens has a refractive index of 1.8. The beam leaving the lens has a diameter of about 50 μm and diverges only slightly. *C* core of fibre. *CL* cladding of fibre. *CA* glass capillary tube. The lens is cemented to the capillary tube.

coupling efficiency is obtained under these conditions. The laser mode and the fibre mode must of course be properly matched, both in cross-section and divergence. In the lowest-order case considered here, the transverse intensity distribution has only one maximum.

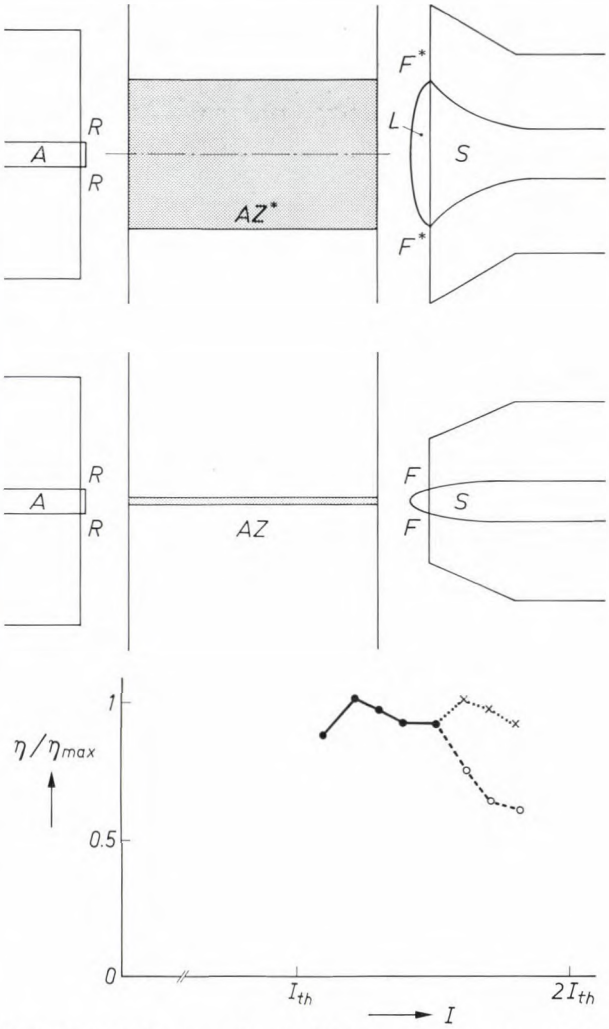


Fig. 6. A GaAs laser can in principle be connected to two single-mode fibres. A tiny semi-transparent mirror *RR* is seated on the core of one fibre, the auxiliary fibre *A*. The other fibre, the signal fibre *S*, has a core end face (elliptical section, horizontal axis *F*F**, vertical axis *FF*) specially adapted to the laser aperture, with a semi-ellipsoidal lens *L* over the core end face. The auxiliary fibre feeds a portion of the generated laser light as a control signal to a circuit that stabilizes the bias current to the laser. The energy in the light pulses is therefore no longer dependent on the temperature in the laser. The light reflected back into the laser by *RR* improves the single-mode operation of the laser and thus has a favourable effect on the coupling efficiency. The graph shows that the single-mode operation of the laser, due to the extra mirror, increases the range of injection-current values with a high coupling efficiency. *AZ* active zone in the laser, vertical section. *AZ** the same, horizontal section. ooo without mirror, ××× with mirror.

[2] P. M. Boers, M. T. Vlaardingerbroek and M. Danielsen, Electronics Letters **11**, 206, 1975.
[3] The characteristics of the various kinds of fibres are described in the article by K. Mouthaan, this issue, p. 178.
[4] W. A. Gambling, D. N. Payne and H. Matsumura, Electronics Letters **10**, 148, 1974.
[5] L. G. Cohen, Bell Syst. tech. J. **51**, 573, 1972.

It will sometimes only be possible to satisfy the requirements that the laser should oscillate in a single mode by making the injection current lower than would be permissible in view of the lifetime of the mirrors [1]. We have succeeded in raising this limit by transmitting a portion of the laser beam back into the active zone with the aid of a tiny mirror; see *fig. 6*. This causes the laser to oscillate in the desired mode even with high injection currents [6]. To ensure correct positioning of this mirror it is made semi-transparent and applied to the end face of the core of a single-mode fibre. (It is correctly positioned when the luminous flux through the fibre is greatest.) The light entering the core of this fibre via the semi-transparent mirror is, of course, very weak and is therefore not taken to the transmission system for use as a communication signal; it can be used as an auxiliary signal for control of the laser bias current as mentioned earlier. The signal to be transmitted is obtained by coupling a second (single-mode) fibre to the *other* reflecting side of the laser. This fibre is flattened at its end so that the local cross-section of the core there is approximately elliptical, with the major axis in the central plane of the active layer. A semi-ellipsoidal lens is fitted to the end face of the core. These measures have the effect of matching the laser aperture and the geometry of the fibre better to each other; the coupling efficiency is improved in this way by a factor of about three.

Summary. A GaAs laser with a double heterojunction can be modulated directly by variation of the injection current. A supply circuit has been devised to deliver a bias current that remains equal to the threshold current independently of temperature. This arrangement greatly reduces the initial delay and keeps the intensity of the light pulses constant (modulation rate up to 500 Mbit/s). A universal coupling mechanism (tolerances about 0.2 μm) allows the relative positions of layer and fibre to be adjusted to give maximum power transfer. In addition to the signal fibre, it is also possible to couple an auxiliary fibre for control of the bias current. A spherical lens can be used with multimode fibres. In single-mode fibres a semi-ellipsoidal lens on the flattened core of the signal fibre increases the coupling efficiency by a factor of 3; a small mirror on the core of the auxiliary fibre increases the range of injection-current values with high coupling efficiency.

[6] G. D. Khoe, Power coupling from junction lasers into single mode optical fibres, IEE Conf. Publn. 132 (1st Eur. Conf. on Optical fibre communication, London 1975), pp. 114-116.

The avalanche photodiode

L. J. M. Bollen, J. J. Goedbloed and E. T. J. M. Smeets

Introduction

If semiconductor photodetectors now available are considered in terms of their usefulness in an optical communication system, the most promising seems to be the avalanche photodiode. In this article we shall consider a number of device-appraisal criteria and see what consequences they have for the manufacture of avalanche photodiodes.

When a voltage is applied across a semiconductor diode in the reverse direction, an electric field of strength E is generated in a region around the P-N junction, since the free charge carriers are drawn away from the junction by the applied voltage. The field gradient dE/dx can be derived from the local doping concentration n_d by using Poisson's equation:

$$\frac{dE}{dx} = \frac{n_d e}{\varepsilon}, \qquad (1)$$

where e is the elementary charge and ε the absolute dielectric constant of the semiconductor material. For a given doping profile the strength of the field and the thickness of the depletion layer can be varied within certain limits by varying the applied voltage; see *fig. 1*.

When a window transparent to the radiation to be detected is made in one of the diode contacts, the radiation is able to penetrate into the diode material. A radiation quantum (photon) that has an energy greater than the energy gap between valence band and conduction band of the semiconductor material can then be absorbed, resulting in the generation of an electron-hole pair. In the depletion layer the electric field separates the electron-hole pair (electrons go to the N region, holes to the P region), and this has the result that a current, the photocurrent, starts to flow in the circuit in which the diode is incorporated.

The situation so far described is what is found in a 'normal' semiconductor photodiode. The avalanche photodiode makes use of the fact that the free charge carriers are accelerated in the electric field and thereby gain kinetic energy. They may gain so much energy that they can ionize lattice atoms with which they collide, giving rise to new electron-hole pairs. In their turn the electrons and holes of these pairs can be accelerated and cause ionization by collision, with the result that a single incoming photon can give rise to an avalanche of electrons and holes. In this case the photocurrent in the external circuit is greater than in the ordinary photodiode described above. The avalanche photodiode is thus a photodiode with internal amplification (multiplication), and may therefore be regarded as a solid-state analogue of the photomultiplier tube.

Ir L. J. M. Bollen, Dr J. J. Goedbloed and Ir E. T. J. M. Smeets are with Philips Research Laboratories, Eindhoven.

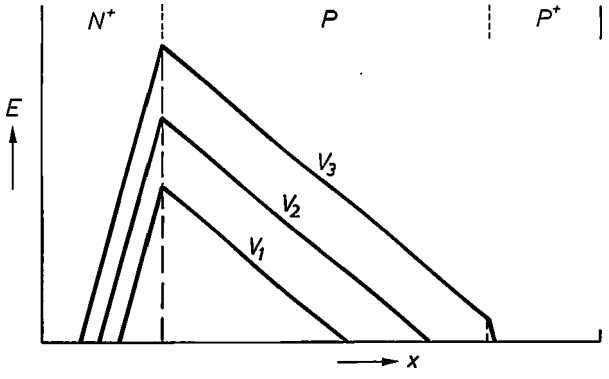


Fig. 1. The field-strength E in an N^+-P-P^+ diode as a function of the depth x below the diode surface for increasing values V_1, V_2 and V_3 of the reverse voltage across the diode.

The main factors that determine the usefulness of a detector in an optical communication system are the following:

— The quantum efficiency, which indicates the proportion of incoming photons detected as a current (without multiplication).

— The response time τ, which determines the maximum modulation frequency of the light for good detection.

— The multiplication factor M of a detector with internal amplification.

— The noise characteristics of the detector.

— Various practical requirements, such as a not unduly high voltage and the avoidance of unwanted side effects.

Some numerical values of M and τ for a number of semiconductor photodetectors are given in *Table I*.

We shall now consider the consequences of the various factors above for the case of an avalanche photodiode, with silicon as the semiconductor material.

Table I. Values of the internal multiplication M and the response time τ for a number of semiconductor radiation detectors.

Detector	M	τ
Photoconductor	10^5	10^{-3} s
P-N diode	1	10^{-11}
P-I-N diode	1	10^{-10}
Phototransistor	10^2	10^{-8}
Avalanche photodiode	10^3	10^{-9}
Field-effect transistor	10^2	10^{-7}

Quantum efficiency

The quantum efficiency is determined primarily by the absorption coefficient of the detection radiation and by the length of the light path in the sensitive part of the diode. The radiation intensity decreases exponentially along this path. The radiation originating from the GaAs semiconductor laser has a wavelength of 0.9 μm. Consequently the energy of the photons is only slightly greater than the band gap in silicon (1.1 eV) and the absorption is low. For radiation of 0.9 μm the absorption coefficient in silicon is about 400 cm^{-1}, which means that the intensity has decreased to 1/e of the initial value over a distance of 25 μm. The fraction of the incoming radiation that is absorbed in the depletion layer is detected in any case, and there is still some detection in zones bordering on the depletion layer. Some of the electrons generated in the P region outside the depletion layer are able to reach the depletion region by diffusion and are then detected, and the same applies to some of the holes from the N^+ region.

To achieve a good quantum efficiency it is therefore desirable to have an active detection region of length about 25 μm in the diode.

Speed of response

Important factors that determine the speed of response are:

— The finite speed of the multiplication process (the charge carriers must pass through a field of a certain strength before they acquire sufficient energy to cause ionization).

— The time that charge carriers generated outside the depletion layer require to reach the depletion region by diffusion.

— The time a charge carrier requires for crossing the depletion layer.

The first factor can be eliminated in practice by making the multiplication zone small and the field-strengths in it very high. Next, since the diffusion always takes place relatively slowly, we must make the diode in such a way that the contribution made by the diffusing charge carriers ('tailing') to the photocurrent is negligible. Finally, if we look at the drift-velocity characteristic of the charge carriers as a function of the electric field-strength we see that at field-strengths greater than about 40 kV/cm in silicon there is no further increase in the drift velocity, which reaches a saturation value of about 10^7 cm/s.

The requirements imposed on quantum efficiency and response time, taken together, result in an even stricter requirement: the diode must have a depletion region about 25 μm long, the part in which the high field-strengths occur must be relatively small, and the field-strength throughout the depletion region must be greater than 40 kV/cm. The time required for crossing the depletion region is then 0.25 ns, which is short enough for optical communication purposes. There must be no tailing effect due to diffusion of charge carriers from outside the depletion region.

Noise characteristics

The generation of electron-hole pairs by the absorption of photons, like the subsequent internal amplification by the multiplication of charge carriers, is a statistical process. It is therefore not surprising that noise plays an important role in the avalanche photodiode. Nevertheless, when the signal-to-noise ratio is considered, this detector comes out well when the other requirements are taken into account. From studies of the noise behaviour of the avalanche photodiode it follows that the α/β ratio, where α is the ionization coefficient of the electrons and β that of the holes, is of essential importance. In silicon it varies from about 10 at a field-strength of about 350 kV/cm to 3 at 500 kV/cm (these are common field-strengths in the multiplication region). The fact that α and β are quite different means that given an otherwise identical distribution of the electrical field in the diode, a better signal-to-noise ratio is found when the charge carriers arriving in the multiplication region are mainly electrons than when the holes are in the majority [1]. This has the result that a new requirement has to be added to those mentioned at the end of the previous section: the part of the depletion region that does not belong to the multiplication region should be situated on only one side of that region and be made of P-type material.

The noise power in a diode of this kind is proportional to $M^2(2 + kM)$, where M is the multiplication by the avalanche process and k is a figure of merit for the diode that depends on the ratio α/β [2]. The proportionality only applies if $k \ll 1$ and $M \gg 1$, conditions that are usually fulfilled in practice. In the total $2M^2 + kM^3$, half of the term $2M^2$ is due to multiplied shot noise during detection, and the other half to shot noise during unilateral multiplication (multiplication involving charge carriers of one kind only). The term

kM^3 is due to non-unilateral multiplication. It is possible to make diodes in which k is about 0.01; a diode of this type can have a gain of 200 before the noise in the internal amplification exceeds the noise in the detection process.

Practical requirements

In addition to the rather fundamental requirement-discussed above, there are various practical requires ments that an avalanche photodiode has to meet. Thus, the dark current, which is due to the electron-hole pairs resulting from the thermal generation in the diode material, must be low. So must the leakage current across the surface of the diode. This is because both currents produce a background and consequently determine the minimum intensity that a light signal must have if it is to be detected.

Another requirement is that the multiplication over the whole active surface of the diode must be as constant as possible. All parts of the diode then have the same amplification and there are no places where premature local breakdown of the diode might occur. From a highly simplified model the relation between the multiplication factor M and the doping concentration n_d is found to be

$$\frac{1}{M}\frac{dM}{M} \approx \frac{dn_d}{n_d}.$$

Thus, with a gain of 100 a relative variation of 10 % in the gain means that the doping concentration must be constant to within 0.1 % over the surface of a diode.

Another desirable feature for future applications of the avalanche photodiode is that the diode voltage should be kept below 200 V. An N^+-P diode with a depletion region 25 μm thick cannot satisfy this requirement. If, however, the P region consists of a relatively thick, very lightly doped layer (the π layer) with a thin, heavily doped P layer on top of it, as indicated by curve *a* in *fig. 2*, the voltage requirement can then be met as well as the other requirements mentioned above. For the voltage requirement a complementary structure, a P^+-N-ν-N^+ diode, could be used, but as we have already noted an N^+-P-π-P^+ diode is preferable because of the better noise characteristics of a diode in which most of the multiplication is due to the electrons.

To make avalanche photodiodes that meet the specifications listed above we started from a substrate of dislocation-free P^+ material. A π layer and then a P layer were first epitaxially deposited on the substrate, and the upper part of the P layer was then converted into N^+ material by ion implantation. In the development of this method of fabrication a fast flying-spot scanning microscope was found to be of great assistance [3]. *Fig. 3* shows a number of photomicrographs, made with this microscope, of mesa diodes fabricated in various ways.

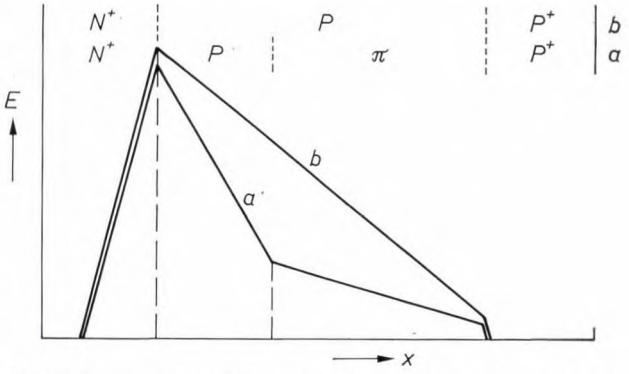


Fig. 2. Curve *a*, the field-strength E in an N^+-P-π-P^+ diode as a function of the depth x below the diode surface. Here the total voltage across the diode (area under the curve) is lower than in an N^+-P-P^+ diode (curve *b*), although in both cases the maximum field-strength and the thickness of the depletion layer are about the same.

We studied two forms of avalanche photodiodes, the mesa diode and the planar diode. Both types, which each have their own advantages and disadvantages, will now be briefly discussed.

[1] P. P. Webb, R. J. McIntyre and J. Conradi, RCA Rev. **35**, 234, 1974.
K. Mouthaan and R. M. Snoeren, Electronics Letters **10**, 118, 1974.
[2] R. J. McIntyre, IEEE Trans. **ED-13**, 164, 1966.
[3] P. M. Boers and L. J. M. Bollen, Philips tech. Rev. **35**, 23, 1975.

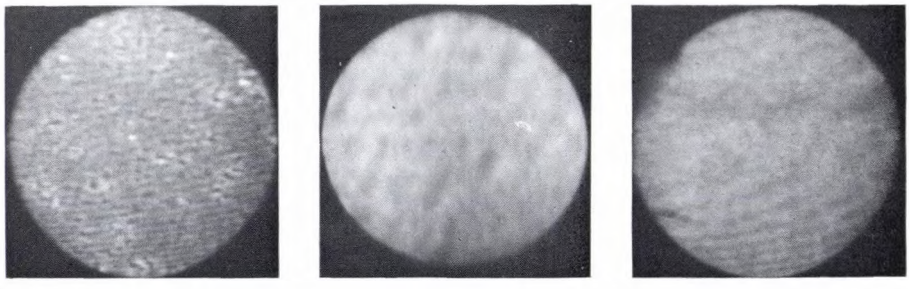


Fig. 3. Pictures obtained with a flying-spot scanning microscope of the photocurrent distribution over the surface of three N^+-P-π-P^+ avalanche photodiodes. The material used was deposited epitaxially on a P^+ substrate. The N^+ upper layer for the left-hand diode was made by diffusion, for the centre one epitaxially and for the right-hand one by ion implantation.

The mesa diode

Fig. 4 shows the cross-section of a mesa diode. This device was made by etching slightly cone-shaped discs from a silicon wafer with the desired doping profile. The substrate used for making the epitaxial layers was previously etched away down to a thickness of about 5 μm and metal contacts were applied to the front and the back. The mesa is given its conical shape by etching with an etchant that attacks the silicon but not the contacts. The effect of this shape is that when the voltage is applied to the diode the depletion layer will extend further at the circumference of the diode than in the middle.

This can be understood as follows. It follows from (1) that in the central part of the diode the distances over which the depletion layer extends on either side of the P-N junction must be inversely proportional to the doping concentrations in the P and N regions. Toward the P-N junction the field-strength on both sides must increase to the maximum value. If a and b are the distances over which the depletion layer extends on both sides, then $aN_1 = bN_2$ (*fig. 5*). Since along the edge there is no material at one side of the P-N junction to provide for this charge neutrality, because of the conical shape, the absence of material there must be compensated by a greater extension of the depletion layer on that side.

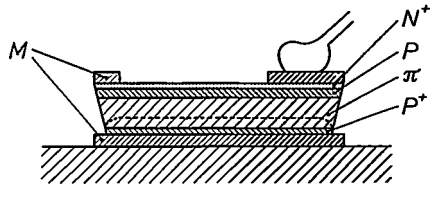


Fig. 4. Cross-section of an N^+-P-π-P^+ mesa diode. The dashed line indicates the boundary of the depletion region in the π layer. The curvature of this boundary at the edge of the diode (greatly exaggerated here) is due to the conical shape of the mesa. M metallization layers.

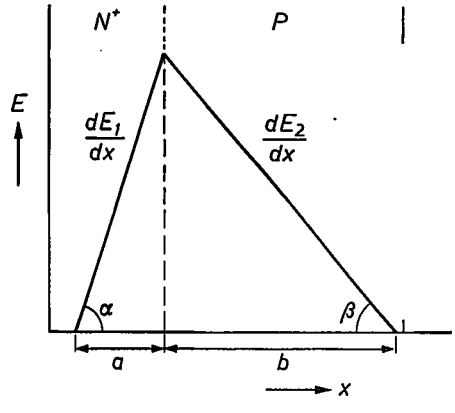


Fig. 5. Field distribution in an N^+-P diode. For the field gradients on either side of the junction the relations: $\tan\alpha = dE_1/dx = N_1e/\varepsilon$ and $\tan\beta = dE_2/dx = N_2e/\varepsilon$ must apply. This means that $aN_1 = bN_2$ for the distances a and b over which the depletion layer extends on both sides of the junction.

Because of the greater length thus available along the edge of the diode for accommodating the same total voltage, the field-strength along the edge will not be so high as in the centre of the diode. This reduces the risk of breakdown along the edge. It is necessary, however, to protect the edges from the effects of water and ions in the atmosphere. The π layer in particular, which is very weakly doped, can easily acquire an N-type skin by adsorption of impurities, giving rise to conducting channels. The edge must therefore be protected by a 'passivating' layer.

The passivation of a mesa diode does not always offer sufficient long-term protection from the formation of N-type channels over the surface of the π region, so that in time the leakage currents may increase. In this respect the use of planar techniques offers better prospects.

Planar diodes

The semiplanar diode

The same material from which the mesa diodes are made, i.e. epitaxial π and P layers on a dislocation-free P^+ substrate, can also be used for etching out a planar diode structure (*fig. 6*). The N^+ layer is again made by ion implantation. The critical region of the diode, where surface breakdowns might occur, is now located at the front, where protection can be applied more easily and more reproducibly than on the mesa diode.

To prevent the extension of conducting channels, which here again may grow across the surface from the N^+ region, a ring of P-type material around the actual diode is left untouched by the etching process.

At the edge of the thin N^+ region the field-strength can become so high that premature breakdown may occur here (*fig. 7*). To avoid such high field-strengths a film of slightly conducting material can be applied. This prevents breakdowns at the edge, but the leakage currents increase to an impermissibly high level. An acceptable solution for this problem is difficult to find.

The planar diode

If we start from material consisting of an epitaxial π layer on a P^+ substrate, a diode of the cross-section illustrated in *fig. 8* can be made by purely planar techniques. The N^+ layer is surrounded here by a relatively thick guard ring with the same type of doping, so that the outer edge of the N^+ region is rounded and high local field-strengths are avoided. In addition the π surface is covered with a silicon-dioxide layer to reduce leakage currents. Here again a ring of P material is applied to stop conducting channels. In this construction both the N^+ layer and P layer are made by implantation in the final stages of the process. This

allows high temperatures to be used in the earlier stages of the process, in forming the oxide layer and in diffusing the N^+ protecting layer, without the risk of damaging any existing thin N^+ and P layers. Both the localization and the concentration of the doping can be accurately and reproducibly controlled by using an implantation process.

The thickness of the P layer is limited by the available implantation energy to 1.5 μm. However, to achieve the desired multiplication it is necessary to have a sufficiently wide region where the field-strength is high enough for avalanche amplification. A consequence of this is that the field gradients are steep and the maximum field-strength high (*fig. 9*), and this has an adverse effect on the noise characteristics of the diode. These drawbacks can be overcome by applying a doping profile like that of *fig. 10* during the implantation of the P region.

In planar diodes the depletion layer in the π region along the circumference of the diode is curved away from the P^+ substrate; this differs from the situation in the mesa diode, for which we saw that the depletion layer curves toward the substrate. It is therefore possible to choose the parameters of the π layer such that

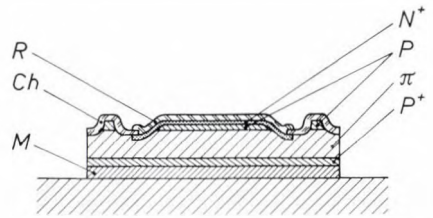


Fig. 6. Cross-section of a semiplanar N^+-P-π-P^+ diode. The surface of the π region is given a protecting layer of slightly conducting material to avoid high local field-strengths. On the actual diode, i.e. on the N^+ material, a special coating is deposited to minimize reflection of the radiation to be detected. M metallization. C conducting layer. R anti-reflection layer.

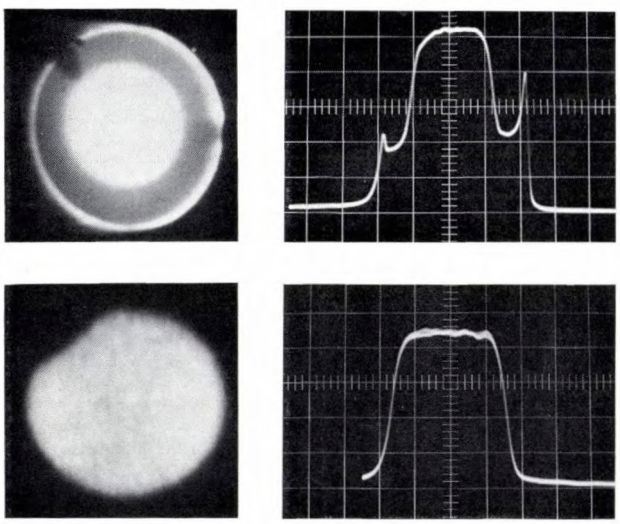


Fig. 7. Effect of a conducting layer on the behaviour of a semiplanar diode. *Above:* a diode without conducting layer: breakdown occurs along the edge. *Below:* a diode with conducting glass layer: no breakdown occurs along the edge. The conduction through the glass increases the dark current of the diode.

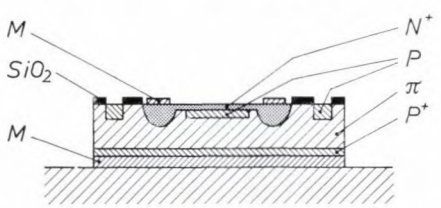


Fig. 8. Cross-section through a planar N^+-P-π-P^+ diode. The edge of the N^+ region is rounded off by an additional N^+ diffusion to reduce the field-strength at the edge and thus prevent premature breakdown at this location. M metallization layers. SiO_2 silicon-dioxide layer.

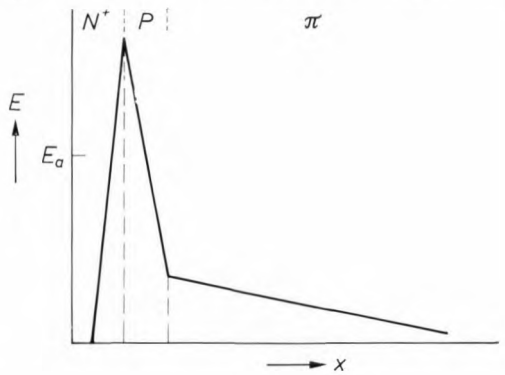


Fig. 9. Effect of the limited thickness of N^+ and P regions in the fabrication of these regions by ion implantation. If the field-strength in the depletion layer is to be raised above the threshold value E_a for avalanche amplification over a sufficiently large distance, the field gradients must be relatively high and consequently the maximum field-strength is high.

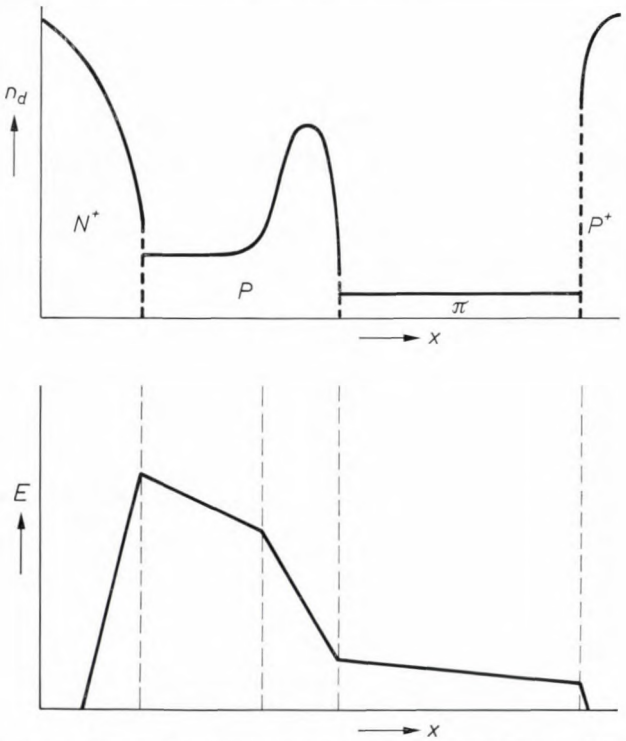


Fig. 10. The net doping concentration n_d and the field-strength E as functions of the depth x below the surface of a diode with a 'low-high-low' structure. In this structure an excessively high field-strength is avoided. The successive regions are not drawn to scale in the x-direction.

at the operating voltage the depletion layer completely reaches the P^+ substrate (*fig. 11*). This means that the field-strength throughout the π region is so high that the charge carriers there acquire the saturated drift velocity, while tailing from the substrate is negligible — which helps to keep the response speed of the diode high. *Fig. 12* shows how the pulse response of an experimental planar N^+-P-π-P^+ diode can be affected by the tailing effect. In one case the voltage is made so low that the depletion region hardly goes beyond the P region; there is then considerable tailing. In the other case the voltage is so high that the whole π region is depleted; there is then very little tailing and the pulse response is greatly improved.

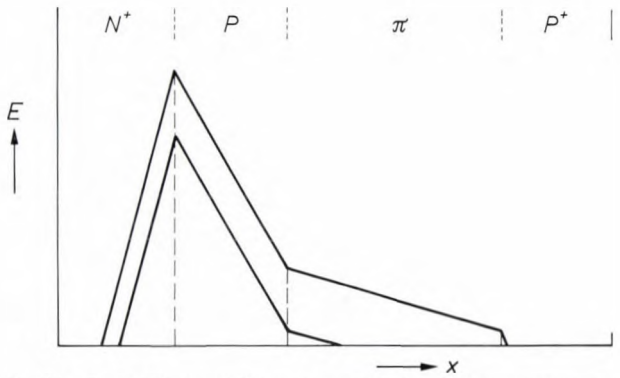


Fig. 11. A relatively slight increase in the voltage across an N^+-P-π-P^+ diode gives rise to such a high field-strength in the whole region that the charge carriers generated there acquire the saturated drift velocity almost immediately.

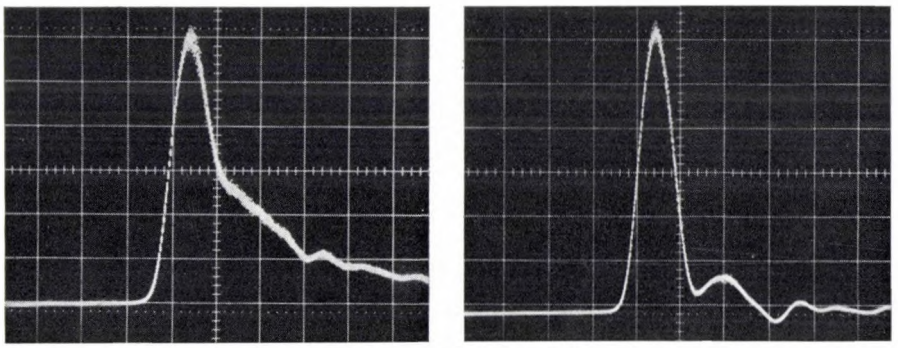


Fig. 12. *Left:* pulse response of an N^+-P-π-P^+ diode in which the depletion region is so narrow that the π region is field-free. *Right:* pulse response of the same diode in which the voltage is so high that the whole π region is depleted.

In the first case the charge carriers generated in the π layer have to reach the depletion region by diffusion. This process is relatively slow and a 'tail' appears on the current pulse produced upon detection of a sharp radiation pulse. In the second case the field-strength in the π region is so high that the charge carriers acquire the saturated drift velocity almost immediately, and there is therefore virtually no tailing. Charge carriers generated in the P^+ layer have such a short life that they have very little effect on the pulse response. The oscillations in the tail are due to the instruments.

Summary. An avalanche photodiode is a photodiode with internal amplification. Charge carriers generated in the diode by photons are accelerated in the internal field of the diode so that on collision with the lattice atoms they generate more charge carriers. This process repeats itself, giving rise to an avalanche of charge carriers for each photon absorbed; the diode is a solid-state analogue of the photomultiplier tube. In a diode built up from an N^+, a P, a π and a P^+ layer the charge carriers taking part in the multiplication process are predominantly electrons, which results in good noise characteristics. Another advantage of this structure is that even with the thick depletion layer necessary to achieve sufficient absorption of the radiation to be detected, the diode voltage remains within reasonable limits. Three versions of the diode are discussed: a mesa diode, a semiplanar diode and a planar diode. The planar diode, made partly by ion implantation in epitaxial material, offers the best prospects.

Characterizing optical fibres; a test bench for pulse dispersion

J.-P. Hazan and L. Jacomme

The extent to which glass fibres can be considered useful as optical waveguides can mainly be derived from a study of their pulse dispersion [1]. A short-duration light pulse propagating in a fibre is subject to a continuous distortion: the amplitude decreases and the 'width' (the duration of the pulse) increases (*fig. 1*). This pulse-broadening or time dispersion is of great importance, since the inverse of the pulse width at the output determines the number of bits of information the fibre can transfer per second: the greater the pulse width the smaller the maximum information rate.

The various kinds of optical fibres and the total dispersion arising in them have already been discussed in the first article of this issue. All that we need do here is to state that in the multimode fibres the most important contributions to the total dispersion are those of the intermodal dispersion and the material dispersion.

The intermodal dispersion — which will receive most attention in this article — is caused by the difference in the propagation times for the various modes. This form of dispersion is strongly dependent on the way in which the light is injected and also on the 'index profile', i.e. the variation of the refractive index in the radial direction in the core of the fibre. Index profiles can be found for which the intermodal dispersion has a minimum; the exact shape of these profiles depends on the wavelength.

The material dispersion arises because the refractive index of the core material in the fibre depends on the wavelength; this dispersion is proportional to $\Delta\lambda$, the spectral width of the light to be guided by the fibre — provided this width is not too great [2]. It can be seen from the above that a considerable number of variables have to be taken into account in determining the dispersion. To investigate the behaviour of an optical fibre in detail it is therefore necessary to have extensive and versatile test equipment available. On the other hand, the measured results cannot be interpreted — and the fibre under test cannot be properly characterized — without the aid of a good theory.

In this article we shall describe a versatile test bench that we have designed ourselves. The measurement procedure will be clarified with the help of a detailed discussion of a practical case. The various problems that occur in the evaluation of an optical fibre will arise naturally in the course of this discussion.

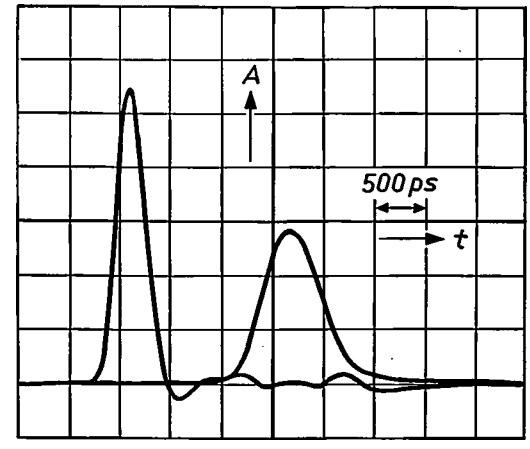


Fig. 1. The short light pulse to the left is the input signal for a multimode fibre with a length of 900 m. The light is injected as a parallel beam. The output signal is shown on the right. The vertical scale shows the amplitude A in arbitrary units, and the horizontal scale shows the time t; 500 ps per division. It can be seen that the pulse becomes wider as it travels through the fibre ('time dispersion') and that the amplitude decreases. The two signals were observed on a 5 GHz sampling oscilloscope.

Fig. 2 is a diagram of the test bench. This is intended for the testing of multimode fibres; if it should be necessary to adapt the bench for single-mode fibres, which could well be of importance in the near future, this would present few difficulties. The bench is built up from components that can in fact be considered to be modular; the light source can be an argon laser or a YAG laser, both with a mode-locked output signal so that the pulses are typically 200 ps FWHM, or a $GaAs/Al_xGa_{1-x}As$ laser or a fast light-emitting diode. The bench also includes a kinematic mount with micrometer adjustment for accurately setting up the light-injection conditions for the fibre. The pulses travel along the fibre and eventually reach the detection and display units. These include a light detector, usually a photodiode of the avalanche type [3] (alternatives are a *P-I-N* photodiode and the 70-ps detector developed at LEP [4]), connected to a sampling oscilloscope or a

Dr J.-P. Hazan and Dr L. Jacomme are with Laboratoires d'Electronique et de Physique Appliquée, Limeil-Brévannes (Val-de-Marne), France.

[1] See for example K. Mouthaan, this issue, p. 178.
[2] D. Gloge, Appl. Optics **10**, 2442, 1971.
[3] K. Mouthaan and R. M. Snoeren, Electronics Lett. **10**, 118, 1974.
[4] J.-P. Hazan, J. Haisma, G. Marie and J. Nussli, IEEE J. QE-6, 744, 1970.

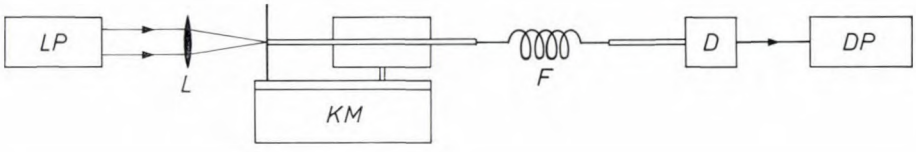
digital-processing oscilloscope (DPO). A fast real-time oscilloscope (5 GHz), developed at LEP [5], can also be used as a display. Finally, it is also possible to connect a spectrum analyser to the system.

The pulse signal emerging from the fibre under test sometimes has a relatively complex shape, such as a multiple-peak structure. The conventional display of the amplitude as a function of time, i.e. the representation in the time domain, is the minimum information that should be available. The DPO and the spectrum analyser offer more possibilities for signal processing and for analysis of the behaviour of the fibre. It is easy with the DPO to obtain the representation of amplitude and phase in the *frequency* domain, i.e. the frequency response of the fibre. The Fourier transforms of the time-domain signals, at both input and output of the fibre, are thus obtainable 'on line'. From these the processor in the DPO can determine the transfer function of the fibre (*fig. 3*).

The DPO can also be used for the convolution of an experimental input signal with a theoretical pulse response; the experimentally measured pulse can then be compared with the theoretical one. In this way theoretical ideas concerning the pulse broadening can be verified.

Experimental investigation of the behaviour of a fibre is of no value if the way in which the light is introduced is not accurately and completely specified. This is because the pulse response is in general strongly dependent on the direction of the incident rays and the coordinates of the point of incidence on the end face of the fibre core.

When a laser is used as the light source it is not in general possible to fully illuminate the whole of the core surface and the total acceptance angle. Various approaches are possible. In one well known method a parallel beam is taken that can illuminate the complete core (*fig. 4a*), and the shape of the pulse and its propagation time are analysed as a function of the direction of incidence. In our test bench this direction is set up by using the kinematic mount. In another method the incident light beam is sharply focused so that the total acceptance angle is completely filled (fig. 4*b*). The beam spot is then much smaller than the core surface. We have extended this method by observing, in addition, the propagation time and pulse shape when the beam spot is displaced radially. We have also developed a 'soft-focus' method, in which a relatively small angle of convergence is associated with a relatively small spot (fig. 4*c*). Provided that the loss in resolution is acceptable, our soft-focus method allows an independent scan to be made by varying either the angle or the radius. With this kind of independent scanning we can more clearly establish the physical origin of the



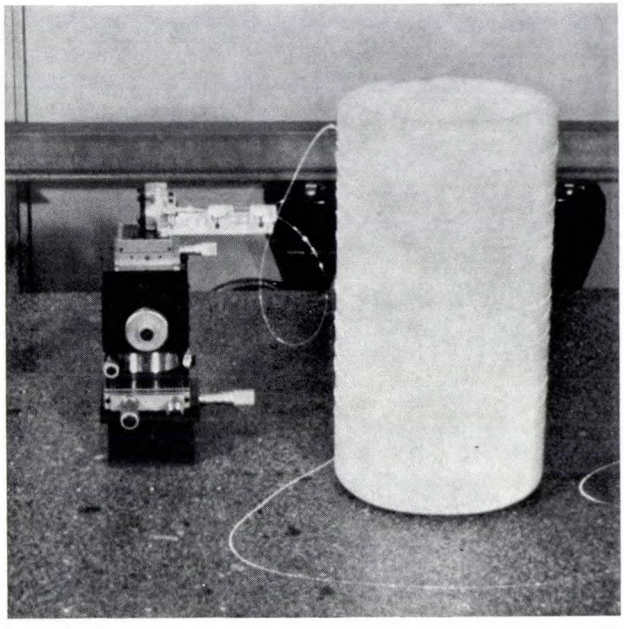


Fig. 2. Block diagram of the test bench for investigating pulse dispersion in optical fibres. The system is built up from modules, so that various kinds of light sources and display devices can be used. *LP* light-pulse source. *L* optics, dependent on the injection conditions. *KM* adjustable kinematic mount. *F* fibre. *D* detector. *DP* display and processing devices. The available light sources are an argon laser ($\lambda = 0.51\ \mu m$), a YAG laser ($\lambda = 1.06\ \mu m$), various GaAs lasers ($\lambda = 0.82\ \mu m$; $0.87\ \mu m$; $0.90\ \mu m$), or an LED. The laser pulses have typically ~ 200 ps full width at half maximum (FWHM).

The photograph shows an adjustable mount for injecting the light into the fibre. The micrometer adjustment ensures that adjustments can be made with lower limits of about 0.01 degrees for angular displacement and about $0.5\ \mu m$ for linear displacements. The read-out from the kinematic mount is digital. The fibre to be tested is wound around the spool next to the kinematic mount.

multiple-peak structure that is observed in certain fibres in the case of parallel-beam injection [6].

Each of the injection methods described here provides a number of basic results that help us to obtain a better understanding of the behaviour of a fibre. In certain applications, however, it is desirable to know the response when both the complete core surface and the total acceptance angle are used. Such a combined result can be derived from a synthesis of a series of results obtained with parallel incident beams at angles of incidence of 0°, 1°, 2°, 3°, etc. The DPO is again well suited to this application.

We have recently succeeded in producing a light source with an increased effective product of radiating surface and emission angle by coupling a sharply focused laser beam into a short length of auxiliary fibre that functions essentially as a 'distributed ray scrambler' (fig. 4*d*). The end surface of the auxiliary fibre, which has a stepped-index profile, consequently behaves as a homogeneous light source that can fully illuminate the core surface and the acceptance angle at the same time [7].

A thorough investigation of a new fibre will in general include the determination of its response at three, or perhaps four, different injection conditions. These experiments will then have to be repeated for various lengths of the fibre, to check whether mode-conversion effects are present, and at various wavelengths. In the quality control of a type of fibre that is already in production, a much simpler test programme is of course sufficient.

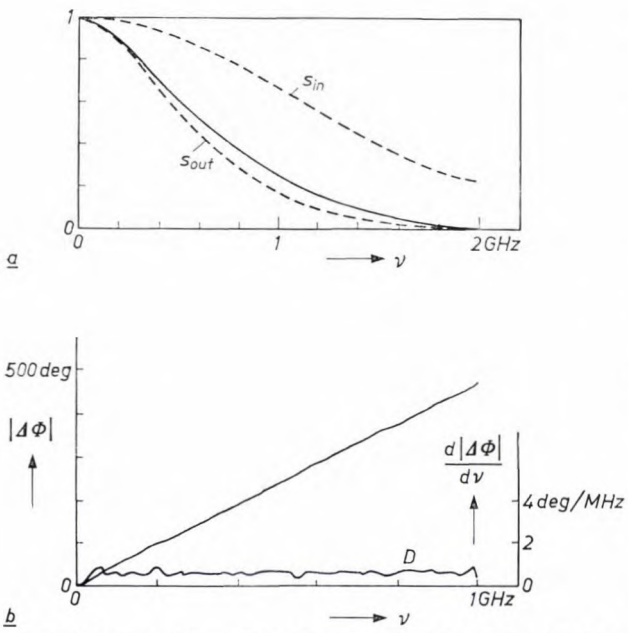


Fig. 3. Frequency-domain representation and transfer function of the pulses shown in fig. 1. *a*) Normalized amplitude (modulus), as a function of the frequency ν, of the input pulse s_{in} and the output pulse s_{out}. The modulus s_{out}/s_{in} of the transfer function is also shown (solid curve). *b*) Phase plotted against frequency. $|\Delta\Phi|$ represents the phase of the transfer function, i.e. the phase difference between the output pulse and the input pulse. The derivative of $|\Delta\Phi|$ with respect to frequency is shown separately (curve *D*); from this derivative the group velocity can be calculated.

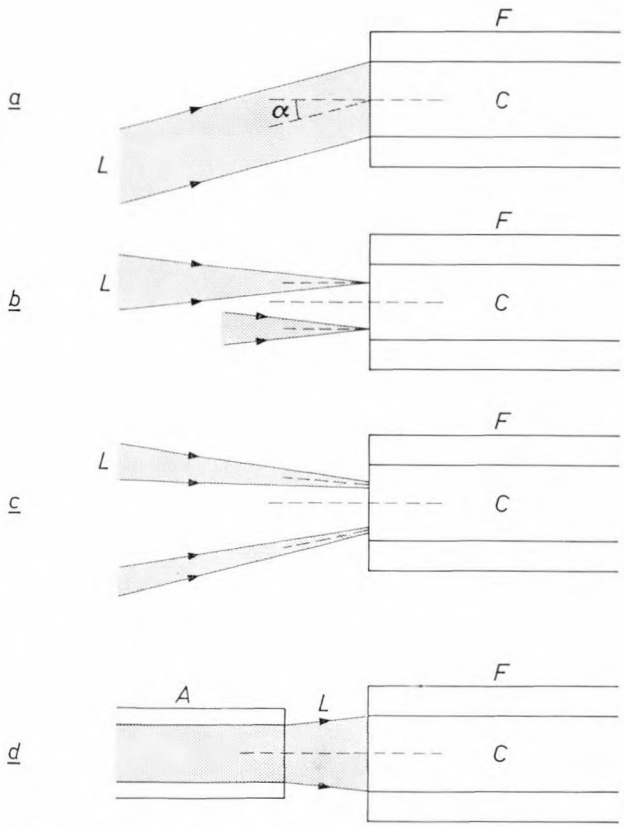


Fig. 4. Four possible ways of injecting light into an optical fibre in the test bench (fig. 2). *F* optical fibre with core *C*. *L* light beam. *a*) Parallel beam. The angle of incidence α can be varied. *b*) Strongly focused beam. The acceptance angle is completely filled; the size of the beam spot on the core surface is very small in relation to the core diameter. The spot can be displaced radially. *c*) 'Soft focus'. The position of the spot and the angle of incidence can both be varied. *d*) 'Ray scrambling'. A large number of rays are continuously mixed in the auxiliary fibre *A*, and the effect of this is that the end surface behaves as a homogeneous light source with a large radiating surface and a large angle of emission.

We have made use of the equipment described here in studying the pulse response of various fibres that have recently become available. One of these tests will be described [6]; the example illustrates some of the practical possibilities of the equipment.

The test fibre had a length of 150 m; its core diameter was between 30 and 34 μm and the acceptance angle was about 24° (for the full cone). The doping profile of the core of the fibre is shown in *fig. 5*; for the kind of fibre used this doping profile gives a fair indication of the index profile. It is clear from the results (fig. 5) that the refractive index of the core

[5] G. Clément and C. Loty, Acta Electronica **16**, 101, 1973.
[6] J.-P. Hazan, L. Jacomme and D. Rossier, Optics Comm. **14**, 368, 1975.
[7] J.-P. Hazan and J. J. Bernard, Short-pulse light sources with large throughputs for optical fibre characterization, to appear shortly.

material had an almost constant value to a radius of about 8 μm; beyond this value the refractive index decreased, arriving eventually at the value for the cladding material.

The first measurements were made with parallel-beam injection giving full illumination of the core. *Fig. 6* is a representation of the observed pulse response of the fibre. The signal includes two peaks; the arrival time of one of them was found to vary with the direction of incidence, whereas the other hardly moved at all — at any rate up to angles of incidence of about 10°. This stationary peak was delayed by about 1.5 ns, if the movable peak at 0° was taken as a time reference. At 9 to 10° the two peaks begin to merge, and at even larger angles a small shift in timing can be observed.

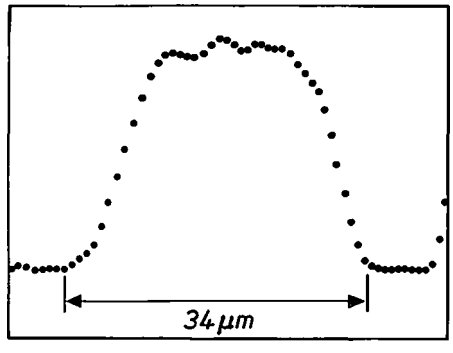


Fig. 5. A concentration profile for the dopant in the core of an experimental fibre. The core diameter is 34 μm. The vertical scale represents the doping concentration, measured with an electron microprobe. This profile gives a fair representation of the radial dependence of the refractive index in the core material.

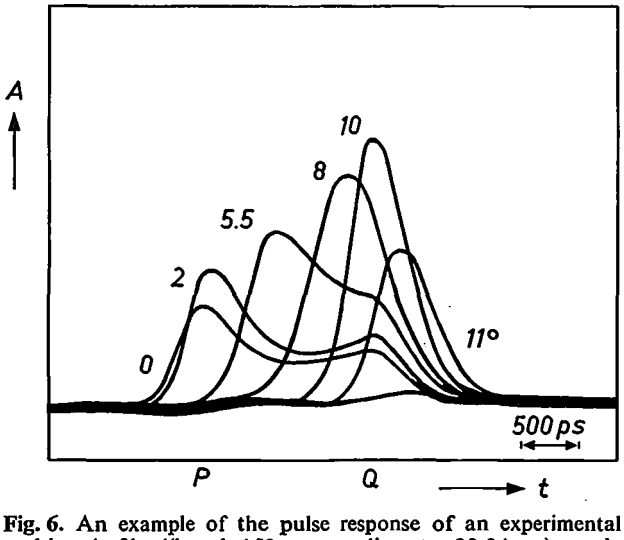


Fig. 6. An example of the pulse response of an experimental multimode fibre (length 150 m, core diameter 30-34 μm), at six different values of the angle of incidence. The light is injected as a parallel beam. The output signal was measured with the sampling oscilloscope (fig. 2). The angle of incidence of the input signal, obliquely incident parallel light, was varied by means of the kinematic mount. Peak *P* moves as the angle of incidence increases, whereas peak *Q* shows hardly any displacement for angles up to 9° to 10°.

The tests on the fibre were completed by a number of measurements using the soft-focus method. These results indicate that the moving peak in the signal originates from light that enters the central part of the fibre core — where the refractive index is almost independent of the radius — and also that the stationary peak is mainly due to light injected at the periphery, in the region where the refractive index has a marked radial gradient [6].

When the investigation referred to above was made, pulse behaviour of this type in fibres had rarely been described, and it was therefore important to find a satisfactory theoretical description.

The theoretical approach that we have developed is a geometrical-optical method based on an analogy with classical mechanics [8]. Its advantage compared with the conventional theory — which is based on a modal description, derived from the electromagnetic wave equation — is that the new theory takes the injection conditions into account and can follow the different rays along the fibre. Every light ray in the core has two characteristic 'constants of motion', comparable to the mechanical quantities of energy and angular momentum for a particle moving in a given path. The output pulse can be determined if the propagation times are computed for a sufficiently large number of rays. The theory can also take account of the 'leaky' rays present, by the use of suitable attenuation coefficients. *Fig. 7* shows the calculated output signals. The most important result is that the new theory does indeed confirm the presence of two peaks that behave like those found experimentally. Obviously the shape assumed for the input signal is an idealized one, where by 'idealized' we mean that the time dependence is assumed to have the form of a delta function and that the light source itself is purely monochromatic. The stationary peak is found to appear 1.65 ns later than the movable one (at incidence angle 0°); this is in reasonable agreement with the experimental value of about 1.5 ns. The results of the calculations depend strongly on the assumed radial dependence of the refractive index. In the calculations for our example we have used a profile with a flat top for half the total width and a gradual slope at the edges (fig. 7*b*). The expression used for the curve for the edges contains a term to the power three as the highest power of the radial coordinate.

Actual index profiles frequently have a rather complex shape and accurate measurements are difficult. This is why it is none too easy to establish a realistic numerical expression for the profiles in the theoretical analysis.

[8] L. Jacomme, Optics Comm. 14, 134, 1975, and Appl. Optics 14, 2578, 1975.

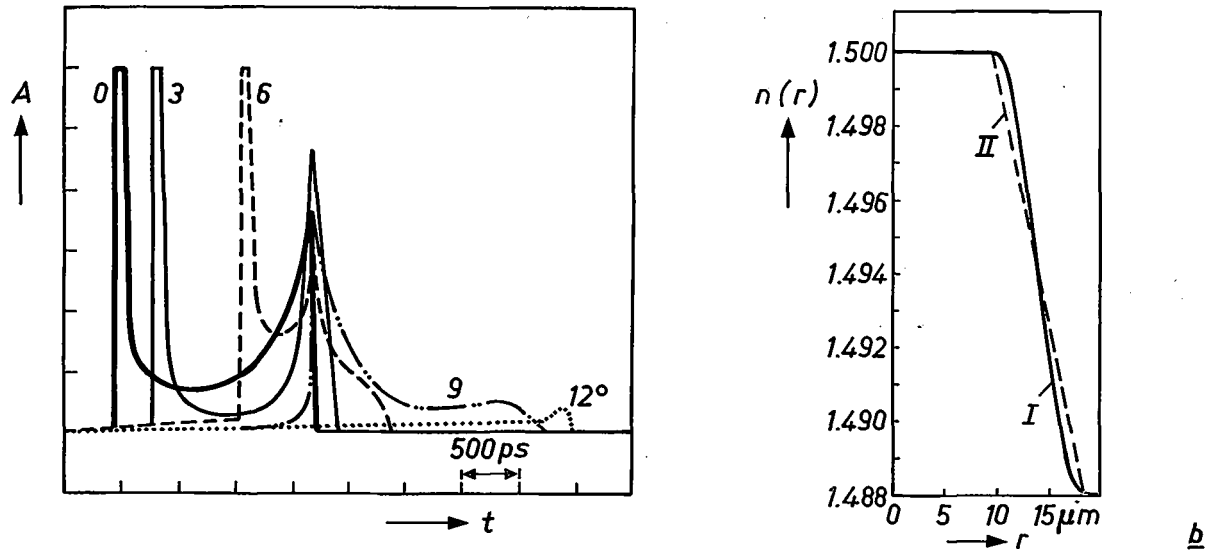


Fig. 7. *a*) The output signal of the fibre, calculated from a new theory, based on a geometrical-optical method [8]. The angle of incidence acts as the parameter. It is assumed that the input pulses are delta functions and that the light source is monochromatic. *b*) Curve *I*, the radial dependence of the refractive index in the core of the fibre; this profile has been used in the calculation of the output signal (*a*). Curve *II*, linear profile used in investigating the sensitivity of the results to small variations in the parameters.

Fig. 8 shows the calculated pulse response when the theoretical behaviour of the fibre, as expressed in fig. 7, is convoluted with a realistic input signal. The convolution has been carried out by the processor of the DPO. Comparison of figs. 6 and 8 shows general agreement between measured and calculated pulses; this indicates that the effects involved here can be explained fairly well by our theory. However, detailed comparison still reveals some slight discrepancies, such as a small displacement of the stationary peak. Such a displacement can be attributed to an error, which could be quite small, in the choice of one of the parameters for the refractive index.

An investigation is still under way at present to find out the extent to which the results are sensitive to small changes in the parameters. If for example the assumed index profile is changed slightly, so that the linear curve *II* of fig. 7 applies, calculations show a pulse

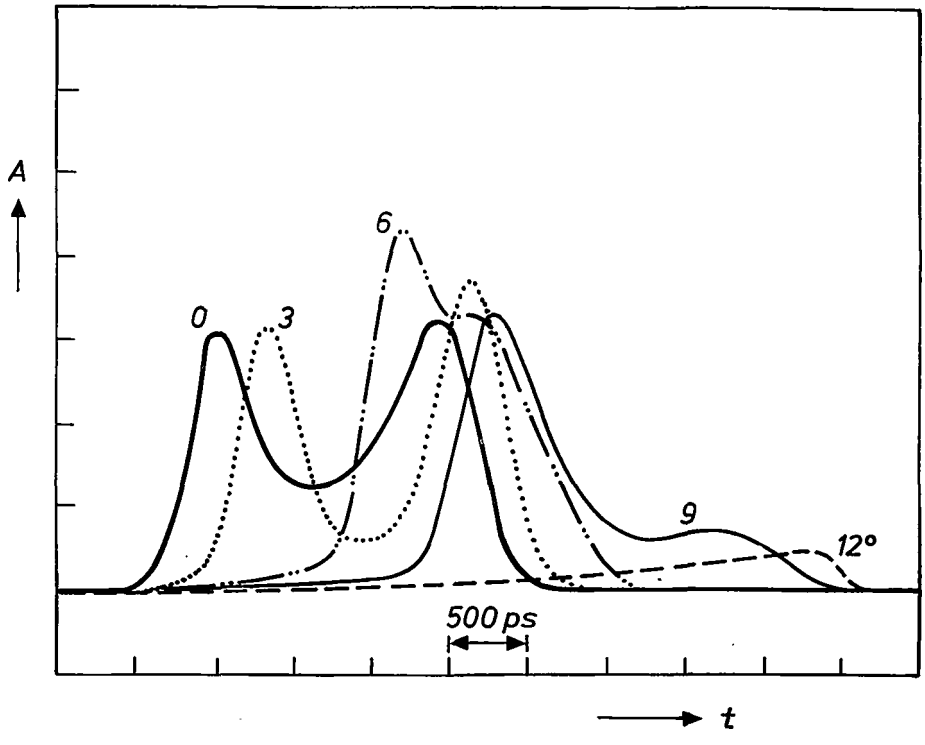


Fig. 8. The output pulses obtained by convoluting the pulses of fig. 7*a* with realistic input signals; a digital-processing oscilloscope was used for the convolution. The observed experimental pulse shapes (fig. 6) correspond fairly well with the ones shown here.

response that still has the double peak, but with the stationary peak about 0.5 ns too early. Another result is that a decrease in the ratio of the flat top to the total width by 6% makes the stationary peak earlier by about 0.16 ns; this corresponds to 10% of the measured time difference between the peaks. Data such as these are of value in deciding the kind of accuracy that is necessary for a reliable comparison of theory and experiment.

In conclusion we should like to emphasize that because of the various parameters involved the pulse-propagation effects in optical fibres can become rather complex. Their detailed understanding requires versatile equipment and a good theoretical insight in performing the necessary experiments. Investigations of this kind will be of value in finalizing the design and setting up the production of optical fibres with large bit-rate capacity.

Summary. A test bench for determining the propagation characteristics of signals in an optical fibre is described. The equipment includes a fast light-pulse source, optical components and an adjustable kinematic mount for the introduction of the light, and also fast display and processing equipment (resolution about 0.1 ns). The system consists of modules and it is very easy to adapt it by altering or changing components. The potential that it offers is illustrated by a discussion of a case in which the pulse propagation takes place in a multimode fibre whose refractive-index profile clearly deviates from the 'stepped' profile. The use of various methods of light injection allows a detailed understanding of the shape of the output signal to be obtained. A geometrical-optical theory, based on an analogy with classical mechanics, predicts pulse shapes that correspond well with the experimental results.

PHILIPS TECHNICAL REVIEW

VOLUME 36, 1976, No. 8

The heat of formation of alloys

A. R. Miedema

Some years ago Philips Technical Review published an article on the alloying behaviour of metals, in which qualitative rules were given. These studies have continued, and have now advanced far enough for the author to be able to give a semi-quantitative description of the known data relating to the energy effects in alloying, at any rate for binary alloys of liquid metals or solid alloys with at least one transition metal. This means that, with the exception of a relatively small group of solid alloys of two non-transition metals, the heat of formation can now be predicted with an accuracy comparable to that of the current experimental methods of measurement. In this article the author explains how the step 'from qualitative to quantitative' is made, and gives many examples to show that knowledge of the heat of formation of alloys can be a great help to the metallurgist in solving a wide variety of practical problems.

Introduction

In chemistry three major branches can be distinguished: organic chemistry, inorganic chemistry and the chemistry of metals. The last of the three has received relatively little attention, especially in education; it can in fact be said to have been somewhat neglected. Many readers may therefore find it rather surprising to learn that it makes sense to draw a sharp dividing line between the chemistry of metals and inorganic chemistry. Very roughly, inorganic chemistry may be defined as the chemistry of compounds in which at least one of the elements comes from the halogen family or the oxygen group. This definition clearly excludes a large number of combinations of elements, which, as can be seen from the periodic system sketched in *fig. 1*, are mainly combinations of two *metals*.

The special nature of combinations of metals appears from the following example. In inorganic chemistry the proportions in which two elements will form a compound can be predicted fairly easily. Thus, Na and Cl will of course form the compounds NaCl; Al and O combine to form the compound Al_2O_3. The essential simplicity of this is related to the generally accepted concept of the 'valency of chemical elements'. We know that in compounds Na is preferentially monovalent, Al is trivalent, Cl is monovalent negative and O divalent. For the formation of compounds from two metallic elements, on the other hand, the valency concept is no longer relevant [1]. Thus, from the elements La and Ni five compounds [2] may be formed, $LaNi_5$, La_2Ni_7, $LaNi_3$, $LaNi_2$ and La_2Ni_3, in which the numerical ratios may be regarded as arbitrary.

The problem of which compounds will be formed and the crystal structures they will have is therefore an intractable one. For a given combination of two metals it is however possible to state definitely whether compounds will be formed or not, how many there will be, and in particular to say something definite about the energy effect that occurs on the formation of a compound from the pure elements.

The heat of formation or heat of reaction is of course of central importance in all chemistry and not just in the chemistry of metals. In the limited case of the halides or chalcogenides (the oxygen group) the heat of

Dr A. R. Miedema is with Philips Research Laboratories, Eindhoven.

[1] An introduction to the subject of intermetallic compounds is given in J. H. N. van Vucht, Philips tech. Rev. 36, 136, 1976.

[2] K. H. J. Buschow and H. H. van Mal, J. less-common Met. 29, 203, 1972.

formation is a well known quantity for many of the possible compounds. This is certainly not true for intermetallic compounds; although a great deal of experimental data has been collected [3], taken together all these data only relate to a small fraction of the number of possible compounds. There are no experimental data at all for the heats of formation and the heats of mixing for more than 90% of the combinations of two metals, particularly those of transition metals.

In this article [4] the intention is to make it clear that, although there are often no experimental data, it is still generally possible to give a reasonably accurate prediction of the heats of the formation of intermetallic compounds and also of quantities such as the heats of solution of metals in liquid metals. These predictions are based on an atomic model for alloys that has been discussed earlier in this journal [5]. Since then a number of refinements have been incorporated in the model, as a result of which, in addition to the more qualitative conclusions as to whether or not two metals will form alloys, a quantitative description of the heat of formation can now be given.

The model and the changes in it since 1973 are described in the next section. To illustrate the results achieved a table is given that presents the predicted heats of formation of all 1:1 compounds of the transition metals.

It has meanwhile been found from experience that a table of heats of formation can be a useful aid to metallurgists. The various examples discussed below will demonstrate this. A remarkable result of the investigation is that certain metals exhibit a behaviour that is different from what would be expected from 'chemical intuition': in this field the familiar inorganic chemical analogies in the behaviour of elements are sometimes found *not* to apply. Some typical examples will be discussed.

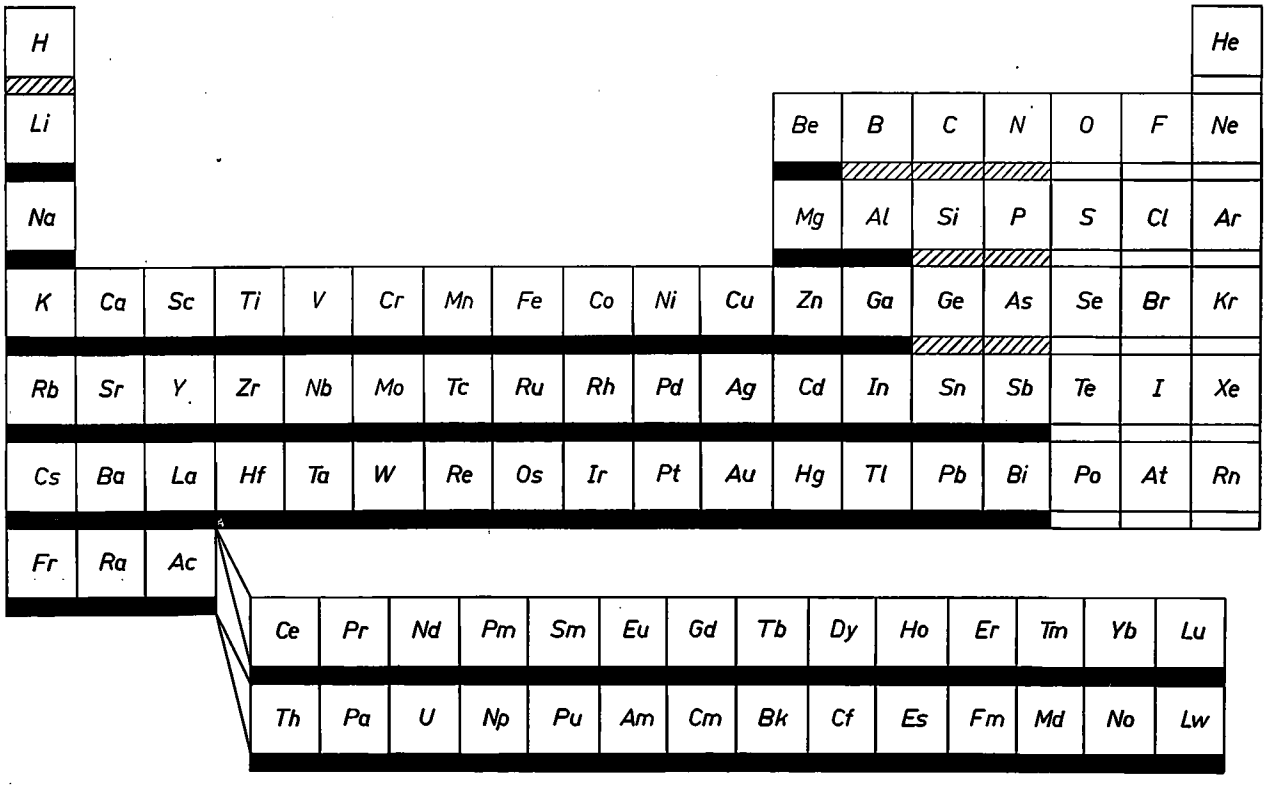


Fig. 1. The elements arranged in accordance with the periodic system. *Black:* metals. *White:* non-metals, which, even in combination with metallic elements, do not usually yield metallic compounds. *Hatched:* borderline cases.

Model for energy effects in alloying; from qualitative to quantitative

The theory of the energy effect that occurs upon the fusion of two metals A and B is based as before on the model illustrated in *fig. 2*. An alloy — shown here with ordered A and B sites, i.e. an intermetallic compound AB — is constructed by taking the atomic cells from the pure metals A and B and placing them with unchanged charge-density distribution in the alloy. Although this changes the shape of the B cells somewhat — otherwise it would not be possible to fill up the whole space — the resultant energy effect is negligible as long as the atomic volume remains the same.

What then remain are contributions to the alloying heats that are related to the fact that the charge distribution does change slightly, i.e. at the interface of the A and B atoms. In the first place there is the effect of the difference in chemical potential for electrons, Φ^*, which existed for the pure metals A and B (contact-potential difference). This difference, which in fact means that electrons have a preference for one of the two metals, leads to a displacement of electrons to places with a more negative potential, and thus to a lowering of the energy of the alloy. This effect, which is always negative and determined by the difference in the

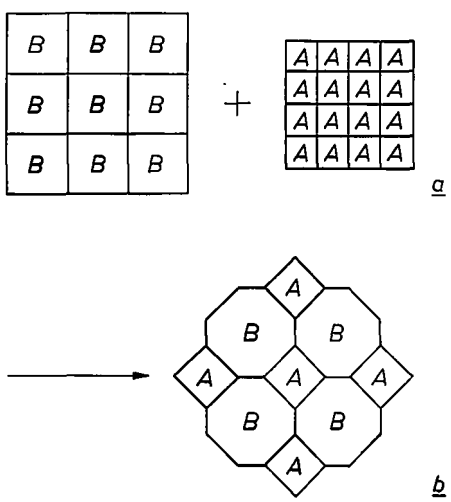


Fig. 2. Atoms in metals (*a*) and alloys (*b*). The alloy AB is considered to be composed of atomic cells that resemble those of the pure metals A and B. In the equalization of the electron chemical potential, electrons are transferred so that the volume of the B atoms increases and that of the A atoms decreases (*c*); the transfer here is from B to A.

chemical potentials $\Delta\Phi^* = \Phi_A^* - \Phi_B^*$, can be approximated by the first term of a series expansion in $\Delta\Phi^*$. This is $-Pe(\Delta\Phi^*)^2$, where P is an empirical constant of dimension V^{-1} and e is the elementary charge.

A second contribution to the heat of formation of the alloy arises from the discontinuity in the electron density that occurs at the interface between dissimilar atoms when atoms A and atoms B are brought together. This difference in electron density Δn_{WS} leads to a positive contribution to the energy of the alloy. In the same way as for the previous contribution this can be written as $Q(\Delta n_{WS})^2$, so that the equation for the formation enthalpy of alloys of two transition metals becomes:

$$\Delta H \propto -Pe(\Delta\Phi^*)^2 + Q(\Delta n_{WS})^2. \qquad (1)$$

As before [5], Q is regarded as a constant for a group of alloys and n_{WS} is the density of electrons at the boundary of the atomic cell. The quantity ΔH is of course determined not only by these two terms, but also by the total number of atoms and the concentration of A and B atoms present.

In the earlier article [5] the validity of equation (1) was verified by comparing the sign of ΔH for combinations of two metals, as found from the characteristics of the phase diagram (see also fig. 3) with the values of $\Delta\Phi^*$ and Δn_{WS} for a given combination. At constant Pe and Q the sign of ΔH is found by comparing the ratio $(\Delta\Phi^*/\Delta n_{WS})$ with $(Q/Pe)^{\frac{1}{2}}$. If $(\Delta\Phi^*/\Delta n_{WS})$ is greater than this new constant, then ΔH is negative; in the opposite case it is positive. These rules have been found to be very reliable. By dividing the alloys into large groups (such as alloys of two transition metals, alloys of transition metals with alkali and alkaline-earth metals or of transition metals with the noble metals Cu, Ag and Au), the value of $(Q/Pe)^{\frac{1}{2}}$ appropriate to each group could be found.

What has changed in the description? We have noted [6] that equation (1) is incomplete in the sense that the constant Q in fact contains a factor $n_{WS}^{-\frac{4}{3}}$. Now for small Δn_{WS} the term $(\Delta n_{WS})^2/n_{WS}^{\frac{4}{3}}$ is equivalent to 9 $(\Delta n_{WS}^{\frac{1}{3}})^2$, as can be seen by differentiating and squaring $\Delta n_{WS}^{\frac{1}{3}}$. Equation (1) thus becomes:

$$\Delta H \propto -Pe(\Delta\Phi^*)^2 + Q_0(\Delta n_{WS}^{\frac{1}{3}})^2. \qquad (2)$$

What exactly is the change? A parameter that was first called Δn_{WS} is now called $\Delta n_{WS}^{\frac{1}{3}}$, but the description in terms of two contributions remains essentially the same. Nevertheless equation (2) is a considerable advance on equation (1). If we repeat the analysis of the sign of ΔH in terms of $\Delta\Phi^*$ and $\Delta n_{WS}^{\frac{1}{3}}$ for various types of alloys, we find that the value of $\Delta\Phi^*/\Delta n_{WS}^{\frac{1}{3}}$ for which ΔH becomes zero — this is equal to $(Q_0/Pe)^{\frac{1}{2}}$ — is identical for alloys of all kinds. In other words, whereas the ratio Q_0/Pe was initially a parameter that had to be empirically determined separately for each group of combinations of two metals, it has not become a kind of natural constant.

This statement is expressed in more detail in *fig. 3*, in which all the data on the sign of ΔH for alloys of two transition metals and those of one transition metal and one of the metals Li, Na, K, Rb, Cs, Ca, Sr, Ba, Cu, Ag and Au are included. The figure is self-explanatory. A straight line gives an almost perfect division between

[3] R. Hultgren, P. D. Desai, D. T. Hawkins, M. Gleiser and K. K. Kelley, Selected values of the thermodynamic properties of binary alloys, Amer. Soc. for Metals, Metals Park, Ohio, 1973.

[4] The predictions of heats of formation given in this article were arrived at in cooperation with Dr F. R. de Boer of the University of Amsterdam and Dr R. Boom of the Hoogovens research laboratories, IJmuiden.

[5] A. R. Miedema, Philips tech. Rev. 33, 149, 1973.

[6] A. R. Miedema, R. Boom and F. R. de Boer, J. less-common Met. 41, 283, 1975, and 46, 67, 1976.

the systems for which ΔH is negative and those for which ΔH is positive. The slope of the straight line corresponds to the value of $(Q_0/Pe)^{\frac{1}{2}}$.

The universal validity of relation (2) is additionally underlined by the results of an analysis of the heats of mixing of liquid metals, published by R. Boom *et al.*[7]. Boom and his colleagues analysed the sign of ΔH for liquid alloys in the manner illustrated in fig. 3 and obtained excellent agreement with the straight line drawn in fig. 3, for liquid alloys of transition metals, of transition metals with the alkali or the noble metals and for alloys of two non-transition metals.

Other changes that have been made in the model since the previous article are less far-reaching. It was already known that a third, negative contribution occurs in the heat of formation of alloys of transition metals with polyvalent non-transition metals (see fig. 1). This contribution, a substantial one, was introduced as an extra constant, to be added to equation (1), which did not depend very much on which of the transition metals combined with which of the p metals. By changing to (2), which, including the extra negative contribution R, becomes

$$\Delta H \propto \{-Pe(\Delta\Phi^*)^2 + Q_0(\Delta n_{WS}^{\frac{1}{3}})^2 - R\}, \quad (3)$$

and again using the fact that $(Q_0Pe)^{\frac{1}{2}}$ is a 'true' constant, it becomes possible to achieve greater accuracy in determining the value of R. *Fig. 4* gives the result of the analysis of the sign of ΔH for combinations of transition metals with p metals. Relation (3) corresponds to a hyperbola: the intersection with the horizontal axis is a measure of the value of R. It appears that this value varies somewhat with the number of p electrons available to the non-transition metal; R is smaller for the divalent than for the tetravalent and pentavalent non-transition metals. The refinement introduced here with regard to the value of R/P obviously leads to a more accurate description and to better predictions of ΔH.

To arrive at a quantitative prediction of ΔH it remains to determine the value of P and the dependence on the atomic concentrations. Details of this will be found in the earlier article [5]. It should be noted that in principle the contact area between dissimilar atoms is measured for the concentration dependence; it is important here whether the A and B atoms differ very considerably in size or not. *Fig. 5* shows the curve of the concentration dependence in the case where A and B are the same size; *fig. 6* gives an example of a marked difference in size.

The values of P and of the parameters Φ^* and $n_{WS}^{\frac{1}{3}}$ of the elements, which differ in a few cases from those given in the previous article, can be found from the literature [6].

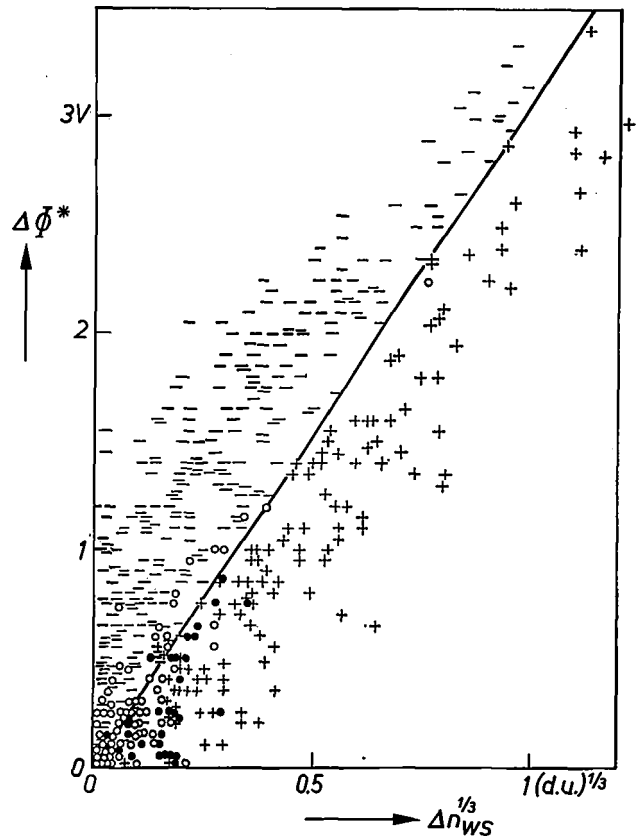


Fig. 3. Test of relation (3) for the solid alloys of two transition metals and those of one transition metal and one of the alkali, alkaline-earth or noble metals. Each binary system gives one point in the diagram, with the coordinates $\Delta\Phi^*$ and $\Delta n_{WS}^{1/3}$.
+ There are no compounds in the system and the solubilities of A in B and of B in A are both less than 10%. It can safely be assumed that ΔH has the positive sign.
— There are compounds or phases with an ordered structure in the system that are stable at low temperatures. It may be assumed that ΔH is negative.
○ There are no compounds or ordered phases in the system, but at least one of the solubilities in the solid state is greater than 10%. It can be postulated that ΔH will not differ much from zero. The figure does indeed show that the open circles mainly occur in the neighbourhood of the origin, which, in the quadratic equation (2), means that ΔH is small.
● As for the open circles, but now the solubility drops to low values at low temperature or there is only incomplete miscibility in the solid state, although both metals have the same crystal structure. The quantity ΔH would be expected to have a small but positive value.

The results of the model description are predictions for ΔH of, for example, ordered intermetallic compounds in a freely chosen concentration ratio. Representative results can be seen in *Table I*, which gives values of ΔH (in kJ/g.at) for the 1:1 compounds of transition metals. In some cases there are no compounds for this concentration ratio, certainly not when ΔH is positive. For other concentrations an estimate of the heat of formation can be made with the aid of Table I and the average concentration dependence of ΔH as depicted in fig. 5. For the heat of solution in dilute alloys, there is a difference corresponding to a

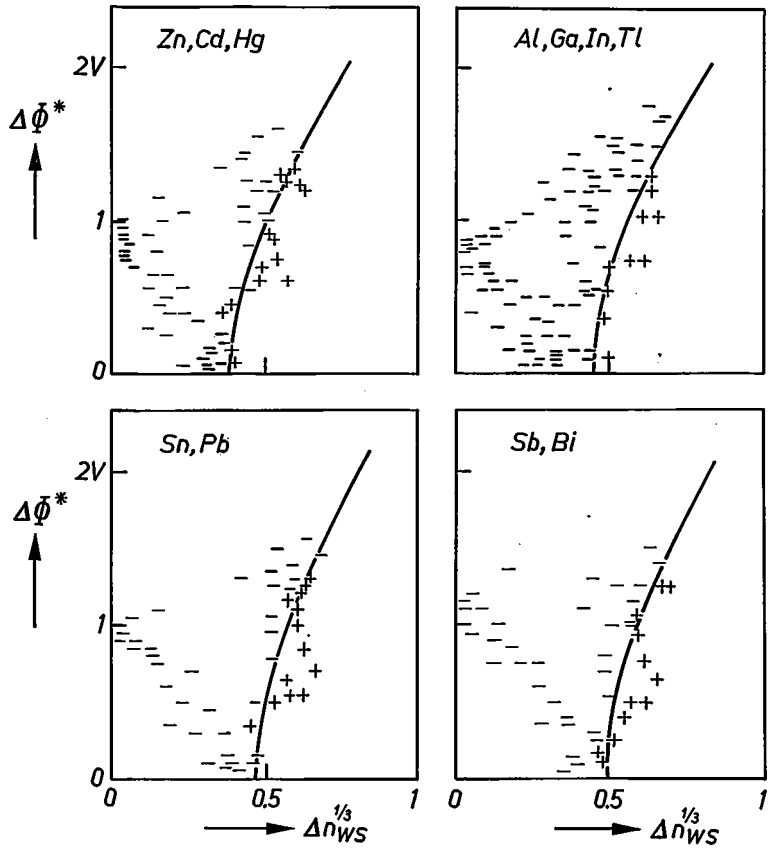


Fig. 4. Determination of the constant R (in fact R/Q_0) in relation (3) for alloys in which a transition metal is combined with a non-transition metal with p electrons. Systems for which ΔH is negative (there are then stable compounds) are separated from those for which ΔH is positive (no compounds, small solubility) by a hyperbola. For all four groups the same value of Q_0/Pe applies: that of fig. 3. It can be seen that the value of R gradually increases with the number of valence electrons of the p metal.

factor of roughly 2.5 between ΔH(AB) and ΔH(A in B), denoted by $\Delta H_A{}^0$.

The calculated values of ΔH as a function of the atomic concentration ratio lie on a continuous line (fig. 5). However, this is only approximately true for experimental values. Energy contributions that depend on the crystal structure are not, it is true, very great, but it is these that are significant when it is a question of which compounds in a system are going to be formed. By way of example, fig. 6 compares experimental and calculated values of Ni-Th, Co-Th and Rh-Th. In all cases the Th atom is considerably larger than the partner atoms, and this has the effect of shifting the minimum in ΔH slightly from the 50/50 composition towards an excess of small atoms. The measure of agreement between calculation and experiment gives an idea of the accuracy that can be expected from the predictions in Table I.

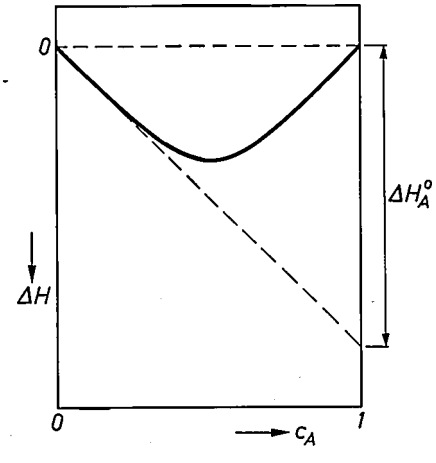


Fig. 5. The solid line indicates the concentration dependence of the heat of formation for a series of ordered compounds in a binary system A-B. $\Delta H_A{}^0$ is the heat of solution of A in solvent B. In this example it is assumed that atoms A and B are the same size.

What do we learn from the heat of formation?

The single number used in Table I for characterizing a combination of two metals nevertheless gives a relatively detailed indication of what will happen on the fusion of these metals. The largest negative values that occur go down to —100 kJ/g.at, and the largest positive values are in the region of +100 kJ/g.at. What does such a value signify? The large negative effect for the combination Pt-Zr, for example, tells us that these elements will react very violently, so violently indeed that the fusion of these metals in considerable quantities might well be dangerous. The specific heat of solids and liquids lies in the region of $3R$ per gram atom (R is the gas constant, equal to 8.3 J). A heat of reaction of 100 kJ/g.at thus implies that the temperature could

[7] R. Boom, F. R. de Boer and A. R. Miedema, J. less-common Met. 45, 237, 1976.

Table I. Model calculation of the heats of formation of intermetallic compounds AB of transition metals (A) with transition metals, alkali and alkaline-earth and noble metals (B). ΔH in units kJ/g.at (1 kcal = 4.19 kJ). Predictions of this form are also available for other atomic concentrations (AB_5, AB_3, AB_2, A_3B_2, A_2B_3, A_2B, A_3B and A_5B), for alloys of transition metals with p metals, including B, C, N, Si and Ge, and for the heats of mixing and heats of solution of all binary liquid alloys. The model cannot be used for solid alloys of two non-transition metals.

	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Y	Zr	Nb	Mo	Tc	Ru	Rh
						3d metals									
Sc		+ 8	+10	+ 1	−12	−17	−42	−55	+ 1	+ 5	+ 24	+15	−57	−64	− 89
Ti	+ 8		− 3	−13	−16	−28	−47	−57	+ 17	− 0	+ 4	− 6	−63	−71	− 86
V	+ 10	− 3		− 3	− 1	−11	−21	−27	+ 24	− 5	− 2	− 0	−32	−37	− 44
Cr	+ 1	− 13	− 3		+ 3	− 2	− 7	−10	+ 15	− 17	− 11	+ 1	−13	−17	− 20
Mn	− 12	− 16	− 1	+ 3		+ 0	− 8	−12	− 2	− 23	− 7	+ 7	−12	−16	− 24
Fe	− 17	− 28	−11	− 2	+ 0		− 1	− 2	− 2	− 36	− 25	− 3	− 5	− 7	− 8
Co	− 42	− 47	−21	− 7	+ 8	− 1		− 0	− 30	− 59	− 38	− 7	− 0	− 1	− 3
Ni	− 55	− 57	−27	−10	−12	− 2	− 0		− 43	− 71	− 46	−11	+ 1	+ 1	− 1
						4d metals									
Y	+ 1	+ 17	+24	+15	− 2	− 2	−30	−43		+ 12	+ 40	+34	−41	−48	− 77
Zr	+ 5	− 0	− 5	−17	−23	−36	−59	−71	+ 12		+ 6	− 8	−79	−88	−107
Nb	+ 24	+ 4	− 2	−11	− 7	−25	−38	−46	+ 40	+ 6		− 9	−56	−63	− 71
Mo	+ 15	− 6	− 0	+ 1	+ 7	− 3	− 7	−11	+ 34	− 8	− 9		−17	−22	− 22
Tc	− 57	− 63	−32	−13	−12	− 5	− 0	+ 1	− 41	− 79	− 56	−17		− 0	+ 1
Ru	− 64	− 71	−37	−17	−16	− 7	− 1	+ 1	− 48	− 88	− 63	−22	− 0		+ 2
Rh	− 89	− 86	−44	−20	−24	− 8	− 3	− 1	− 77	−107	− 71	−22	+ 1	+ 2	
Pd	−127	−108	−53	−22	−34	− 6	− 2	− 0	−121	−138	− 83	−22	+ 6	+ 9	+ 3
						5d metals									
La	+ 11	+ 38	+50	+43	+19	+28	− 3	−17	+ 5	+ 31	+ 71	+70	− 7	−14	− 48
Hf	+ 6	+ 0	− 3	−13	−18	−30	−51	−62	+ 15	− 0	+ 6	− 5	−69	−77	− 95
Ta	+ 24	+ 3	− 2	−10	− 6	−22	−35	−43	+ 40	+ 5	− 0	− 7	−52	−59	− 67
W	+ 14	− 9	− 1	+ 1	+ 9	− 0	− 2	− 5	+ 34	− 13	− 13	− 0	−10	−15	− 14
Re	− 54	− 65	−34	−15	−12	− 6	+ 1	+ 3	− 36	− 80	− 60	−20	+ 1	+ 0	+ 4
Os	− 57	− 67	−35	−16	−13	− 6	+ 0	+ 2	− 40	− 82	− 61	−21	+ 0	+ 0	+ 3
Ir	− 91	− 93	−52	−27	−27	−13	− 5	− 2	− 76	−114	− 82	−32	− 2	− 1	+ 1
Pt	−131	−122	−68	−36	−43	−20	−11	− 7	−121	−151	−104	−42	− 5	− 2	− 2
						actinides and noble metals									
Th	− 0	+ 9	+12	+ 3	−11	−15	−41	−54	+ 2	+ 5	+ 26	+18	−56	−64	− 90
U	+ 7	− 3	+ 1	− 3	− 5	−14	−28	−36	+ 20	− 7	+ 1	+ 2	−42	−48	− 60
Pu	− 1	− 1	+ 6	+ 2	− 5	− 9	−27	−37	+ 6	− 6	+ 11	+12	−38	−44	− 62
Cu	− 43	− 27	+ 6	+20	+ 5	+22	+13	+10	− 40	− 43	− 3	+29	+18	+17	+ 2
Ag	− 50	− 21	+19	+36	+15	+39	+26	+22	− 50	− 43	+ 11	+50	+35	+34	+ 14
Au	−111	− 84	−29	− 0	−17	+12	+11	+11	−109	−115	− 54	+ 5	+21	+23	+ 11

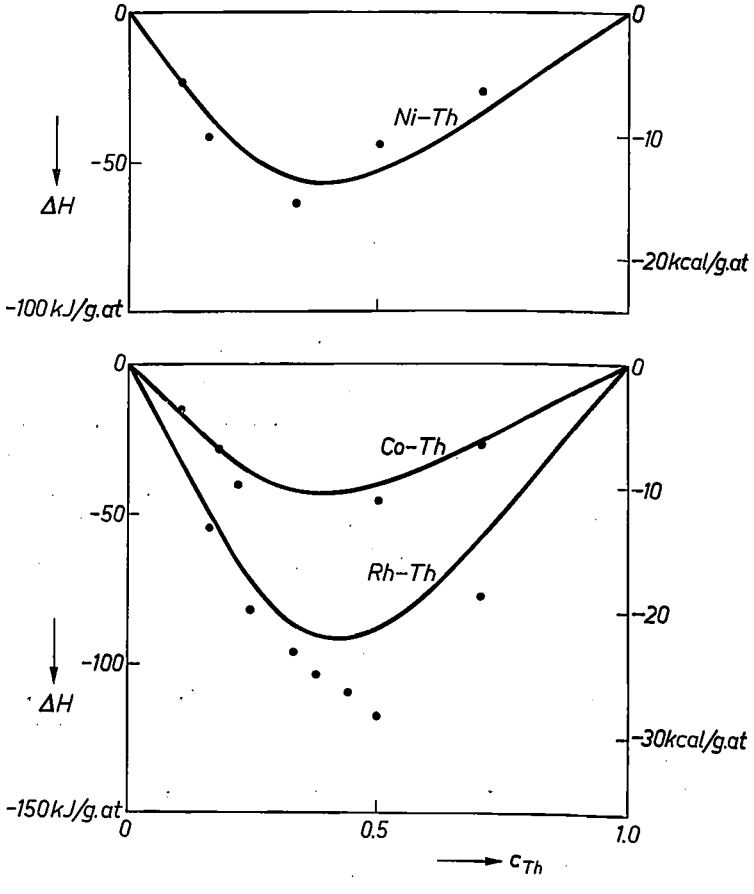


Fig. 6. Concentration dependence of the heat of formation ΔH for the systems Ni-Th, Co-Th and Rh-Th. (The experimental data are due to W. H. Skelton *et al.*[8] and to M. Murabayashi and H. Kleykamp [9].) The curves give the results of the model calculations. The minimum in ΔH does not lie at the concentration 0.50, as in fig. 4, because the atomic volume of Th (19.8 cm³) is much greater than that of Ni (6.6 cm³), Co (6.7 cm³) and Rh (8.3 cm³).

La	Hf	Ta	W	Re	Os	Ir	Pt	Th	U	Pu	Cu	Ag	Au	Li	Na	K	Ca	Sr	Ba	
+11	+ 6	+24	+14	−54	−57	− 91	−131	− 0	+ 7	− 1	−43	−50	−111	+18	+ 51	+ 81	+25	+ 35	+ 39	Sc
+38	+ 0	+ 3	− 9	−65	−67	− 93	−122	+ 9	− 3	− 1	−27	−21	− 84	+44	+ 87	+113	+53	+ 65	+ 69	Ti
+50	− 3	− 2	− 1	−34	−35	− 52	− 68	+ 12	+ 1	+ 6	+ 6	+19	− 29	+55	+102	+122	+60	+ 71	+ 74	V
+43	− 13	−10	+ 1	−15	−16	− 27	− 36	+ 3	− 3	+ 2	+20	+36	− 0	+50	+ 97	+113	+51	+ 61	+ 64	Cr
+19	− 18	− 6	+ 9	−12	−13	− 27	− 43	− 11	− 5	− 5	+ 5	+15	− 17	+29	+ 68	+ 83	+26	+ 35	+ 37	Mn
+28	− 30	−22	− 0	− 6	− 6	− 13	− 20	− 15	−14	− 9	+22	+39	+ 12	+38	+ 86	+101	+34	+ 45	+ 47	Fe
− 3	− 51	−35	− 2	+ 1	+ 0	− 5	− 11	− 41	−28	−27	+13	+26	+ 11	+12	+ 55	+ 67	+ 3	+ 12	+ 14	Co
−17	− 62	−43	− 5	+ 3	+ 2	− 2	− 7	− 54	−36	−37	+10	+22	+ 11	+ 1	+ 44	+ 56	−10	− 1	+ 0	Ni
+ 5	+ 15	+40	+34	−36	−40	− 76	−121	+ 2	+20	+ 6	−40	−50	−109	+11	+ 41	+ 72	+16	+ 25	+ 28	Y
+31	− 0	+ 5	−13	−80	−82	−114	−151	+ 5	− 7	− 6	−43	−43	−115	+38	+ 81	+115	+50	+ 63	+ 68	Zr
+71	+ 6	− 0	−13	−60	−61	− 82	−104	+ 26	+ 1	+11	− 3	+11	− 54	+72	+128	+158	+85	+100	+105	Nb
+70	− 5	− 7	− 0	−20	−21	− 32	− 42	+ 18	+ 2	+12	+29	+50	+ 5	+72	+131	+155	+78	+ 92	+ 96	Mo
− 7	− 69	−52	−10	+ 1	+ 0	− 2	− 5	− 56	−42	−38	+18	+35	+ 21	+12	+ 67	+ 83	+ 2	+ 13	+ 15	Tc
−14	− 77	−59	−15	+ 0	+ 0	− 1	− 2	− 64	−48	−44	+17	+34	+ 23	+ 8	+ 62	+ 77	− 5	+ 6	+ 8	Ru
−48	− 95	−67	−14	+ 4	+ 3	+ 1	− 2	− 90	−60	−62	+ 2	+14	+ 11	−20	+ 27	+ 40	−39	− 30	− 28	Rh
−98	−122	−78	−10	+14	+12	+ 9	+ 3	−130	−77	−86	−14	−10	+ 0	−59	− 21	− 11	−89	− 82	− 82	Pd
	+ 36	+71	+73	+ 2	− 4	− 42	− 90	+ 13	+48	+24	−29	−43	− 92	+ 2	+ 20	+ 43	+ 2	+ 6	+ 8	La
+36		+ 5	− 9	−70	−73	−102	−136	+ 7	− 4	− 3	−34	−31	− 99	+41	+ 87	+119	+53	+ 66	+ 71	Hf
+71	+ 5		−11	−56	−57	− 78	− 99	+ 26	+ 2	+11	− 0	+15	− 49	+72	+129	+159	+85	+100	+105	Ta
+73	− 9	−11		−14	−15	− 24	− 30	+ 17	+ 1	+12	+36	+60	+ 17	+74	+138	+162	+81	+ 95	+ 99	W
+ 2	− 70	−56	−14		+ 0	− 1	− 0	− 53	−43	−37	+24	+45	+ 30	+21	+ 80	+ 98	+11	+ 23	+ 26	Re
− 4	− 73	−57	−15	+ 0		− 1	− 1	− 57	−44	−39	+21	+41	+ 27	+16	+ 73	+ 90	+ 6	+ 17	+ 20	Os
−42	−102	−78	−24	− 1	− 1		+ 1	− 93	−67	−65	+ 7	+23	+ 19	−13	+ 40	+ 55	−31	− 21	− 19	Ir
−90	−136	−99	−30	− 0	− 1	+ 1		−136	−92	−94	−10	− 0	+ 7	−48	− 1	+ 11	−77	− 68	− 67	Pt
+13	+ 7	+26	+17	−53	−57	− 93	−136		+ 8	− 0	−42	−49	−115	+20	+ 58	+ 93	+28	+ 40	+ 44	Th
+48	− 4	+ 2	+ 1	−43	−44	− 67	− 92	+ 8		+ 3	− 6	+ 5	− 53	+51	+107	+138	+61	+ 75	+ 80	U
+24	− 3	+11	+12	−37	−39	− 65	− 94	− 0	+ 3		−18	−15	− 67	+31	+ 71	+ 96	+35	+ 45	+ 49	Pu
− 29	− 34	− 0	+36	+24	+21	+ 7	− 10	− 42	− 6	−18		+ 2	− 9	−15	+ 13	+ 22	−27	− 22	− 22	Cu
− 43	− 31	+15	+60	+45	+41	+ 23	− 0	− 49	+ 5	−15	+ 2		− 7	−26	− 0	+ 8	−44	− 41	− 42	Ag
− 92	− 99	−49	+17	+30	+27	+ 19	+ 7	−115	−53	−67	− 9	− 7		−55	− 21	− 13	−86	− 82	− 82	Au

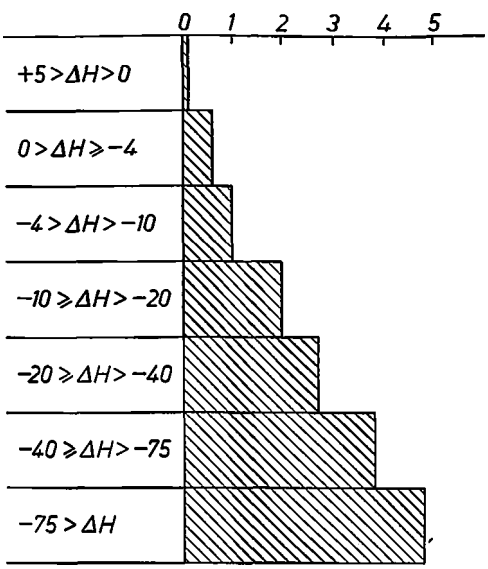


Fig. 7. Connection between the value of the heat of formation of the 1:1 compounds of transition metals, as given in Table I, and the number of ordered phases shown in the associated system. The solid solutions of the two metals are not taken into account here; when ΔH is close to zero, a large solubility is generally found in addition to possible ordered phases, except where the two metals differ substantially in atomic volume.

rise by about 4000 °C if the heat of reaction were not removed. Since the metals are generally melted first to start the reaction, the temperature in a case like this will very quickly reach the boiling point of the metals.

If ΔH goes strongly negative there is a relatively high probability that a number of ordered alloys (compounds) will exist in the relevant system. The energy effects *not* included in the model, which are connected with the crystal structure or for example with the number of conduction electrons per formula unit, have less influence the more the terms that do form part of the model have as their outcome a greater ΔH. In such cases ΔH as a function of the concentration will also tend more to follow a smooth curve (as in fig. 5), so that more compounds can be stable at the same time. In *fig. 7* this correlation is demonstrated for all the binary combinations contained in Table I. The quantity counted is the number of stable compounds occurring

[8] W. H. Skelton, N. J. Magnani and J. F. Smith, Met. Trans. 1, 1833, 1970, and 2, 473, 1971.
[9] M. Murabayashi and H. Kleykamp, J. less-common Met. 39, 235, 1975.

within a binary system, at any rate where the information on the phase diagram is sufficient.

A large positive value, as found in the combination Cr-Na, not only indicates that there cannot possibly be any compounds but also says something about the mutual solubility of the two metals, in this case, for example, the solubility of chromium in liquid sodium. The concentration c of chromium in the solution is given by:

$$c_{Cr} = \exp(-\Delta H_{Cr}^0/RT). \quad (4)$$

With $\Delta H_{Cr}^0 = +260$ kJ (according to fig. 5 the heat of solution is 2.6 times the heat of formation from the table) the solubility at a temperature of say 1000 K is equal to exp(—30), an absolutely negligible amount. For a cooling circuit with liquid sodium, enclosed in stainless steel — as in nuclear reactors — this means that if chromium is after all carried into the sodium it will not be because of metallic chromium dissolving in sodium. In this particular example it appears that the oxygen, which is always present in the sodium, gives rise to the formation of a complex of Na, Cr and oxygen.

For the smaller positive values of ΔH in Table I some sort of prediction about the miscibility of liquids (over the whole concentration region or not) can be made from the numerical value. If ΔH is greater than +20 kJ/g.at, then complete miscibility of the liquid metals is unlikely even at 1000 °C.

Heats of formation are important in the preparation of intermetallic compounds. An example is the preparation of $GdNi_5$, an interesting material for the absorption of hydrogen. It is the practice to make compounds of this type by reducing Gd_2O_3 with Ca in the presence of Ni, the initial consideration being that CaO is more stable than Gd_2O_3:

$$Gd_2O_3 + 3\,Ca + 10\,Ni \rightarrow 2GdNi_5 + 3\,CaO.$$

The data for this reaction at 1000 °K are: $\Delta G(Gd_2O_3) = -1530$ kJ/mol; $\Delta G(3\,CaO) = -1570$ kJ/mol, from which it follows that Ca can reduce the Gd_2O_3 [10].

In this case the heat of formation of $GdNi_5$ is disregarded. In fact it is sufficient for $GdNi_5 + 3$ CaO together to be more stable than Gd_2O_3. This is how a reaction with Ba works (advantage: there are no Ba-Ni compounds):

$$Gd_2O_3 + 3\,Ba + 10\,Ni \rightarrow 2\,GdNi_5 + 3\,BaO.$$

In this reaction [11] 3 BaO yields a smaller formation enthalpy than Gd_2O_3, but the difference is compensated by $\Delta H(2\,GdNi_5)$. Data: $\Delta G(Gd_2O_3) = -1530$; $\Delta G(3\,BaO) = -1385$; $\Delta G(2\,GdNi_5) = -318$, all expressed in kJ per formula unit [12].

In this example, taking the heat of formation of the intermetallic compound into account is no more than a correction, and is just sufficient. In view of the high values of ΔH in Table I, this term need not always be a correction. Thus, because of its extreme heat of formation (—420 kJ/mol) the compound $ZrPt_3$ can be obtained simply by heating ZrO_2 with Pt in a hydrogen atmosphere [13]:

$$ZrO_2 + 3\,Pt + 2\,H_2(gas) \rightarrow ZrPt_3 + 2\,H_2O(gas).$$

Here we may wonder to what extent the predictions contained in Table I are obvious and to what extent they might not be expected by the chemist, familiar with the regularities of the periodic system. It is interesting here to point to the markedly different behaviour of silver and gold. When the ΔH values in Table I are considered, it can be seen that for the metals V, Cr, Fe, Nb, Ta, U, Cu and K the value of ΔH is positive for compounds with silver and negative for compounds with gold. The experimental behaviour is in agreement with this. In all these cases gold forms ordered alloys, whereas silver does not, and in some cases gives phase separation in the liquid.

Similar unexpected behaviour is found (experimentally and in Table I) for the combinations of Y or La with either Mn, Fe or Re, for Sr or Ba with Ni and for the alkali metals with Ir and Pt. The first element of the rare-earth series (La) does not alloy with Mn, Fe and Re, whereas the elements in the middle of the series, such as Gd (to be regarded as identical with Y within the model) do form compounds with these three partners. Sr forms compounds with nickel but Ba does not. Of the alkali metals only Li alloys with Ir, whereas in combination with Pt alloys have been made of Li and Na, but not of K, Rb and Cs. It is typical of the chemistry of metals that the chemical elements belonging to the same column of the periodic system do not necessarily behave in an analogous way.

As can be seen from *Table II*, examples of this type are also to be found among the combinations of transition metals with non-transition metals.

The valency of Eu and Yb in alloys

The calculation of ΔH can be applied for determining the valency of the rare-earth metals Eu and Yb in alloys (valency being the number of electrons outside the ionic nucleus that can be counted among the conduction electrons). As a rule the elements of the rare-earth series in the solid state have three conduction electrons (third column in fig. 1). Exceptions to this rule are Eu and Yb. As a normal rare-earth element Eu should have, in addition to three conduction electrons, another six electrons in the 4f shell which, though not completely filled, does lie in the ionic nucleus. Now

with seven electrons the 4f shell is half full; a half-filled shell presents a favourable atomic configuration, so that elementary Eu gets only two conduction electrons and seven electrons in the 4f shell. In the same way Yb, which would normally have thirteen 4f electrons, gets fourteen of them so that the 4f shell is completely filled. With two conduction electrons Eu and Yb really belong to the Ca, Sr and Ba group. This also appears from the molar volume: in the rare-earth elements this gradually decreases from 22.5 cm^3 for La via 19.9 cm^3 for Gd to 17.8 cm^3 for Lu (known as the lanthanide contraction), but the volume of Eu stands out with 28.9 cm^3 (Yb: 24.9 cm^3). Thus Eu and Yb have more resemblance to Ca (26.2 cm^3). Although Eu and Yb are divalent as metallic elements, this does not of course mean that the metal cannot exist with three valence electrons. This form simply has a higher energy. The corresponding energy gap can be found with the aid of the heats of formation of, for example, oxides or chlorides in which Eu (or Yb) has apparently become a normal member of the rare-earth elements. Thus, Eu with oxygen forms in addition to EuO the compound Eu_2O_3, by analogy with La_2O_3, Pr_2O_3, Nd_2O_3, Sm_2O_3 or Gd_2O_3. The values of ΔH in the formation of these five compounds from metal and oxygen are —1750, —1820, —1805, —1820 and —1820 kJ/mol respectively. For Eu_2O_3 the value of ΔH is clearly less negative, being —1660 kcal/mol. The formation of Eu_2O_3 can be thought of as taking place in two stages: first the trivalent Eu metal is made from the Ca-like Eu; this Eu metal fits into the series of rare-earth elements, and only then is the oxide made from the metal. For this second process the heat effect will be about —1820 kJ per mole of Eu_2O_3, from which it follows that in the first stage about 160 kJ was used to change the valency of 2 moles of Eu. By repeating this type of analysis for various comparable compounds of Eu and averaging the result, the energy difference between the two modifications of Eu [14] can be found to be equal to 95 kJ/mol. For Yb metal this heat of conversion is about 33 kJ/mol.

What happens in alloys? Divalent Eu reacts like Ca, whereas Eu with three conduction electrons is to be found in column Y of Table I. If the heat of formation of a compound YM of Y with the metal M is much more strongly negative than that of the corresponding compound of the metal M with Ca, then Eu will be trivalent in the alloy. For this to be the case, the difference (taken per mole of Eu) must exceed the heat of conversion of 95 kJ/mol mentioned earlier. In *fig. 8* it can be seen that this holds good for compounds in which the valency of Eu is known from experiment and for which the difference in heat of formation has been calculated from the model. The valency of Eu can be derived experimentally from the lattice parameter of the compound (Eu of valency two has a high atomic volume), from the magnetic behaviour (only Eu with seven 4f electrons has an atomic magnetic moment), and from measurements of the Mössbauer effect (which determine the electron density at the atomic nucleus).

In fig. 8 the valency of one of the compounds is given as 2.5. This refers to $EuPt_2$, for which different investigators have given different lattice-parameter values[15].

Table II. Calculated values for the heats of formation (in kJ/g.at) of the 1:1 compounds of the 3d metals (first column) with a selection of non-transition metals (first row). Note that Zn and Cd, In and Tl, Sn and Pb, or Sb and Bi show the opposite sign in a number of cases; this has been confirmed by experiment.

	Zn	Cd	Al	Ga	In	Tl	Sn	Pb	Sb	Bi
Sc	−64	−69	−86	−87	−78	−74	−104	−96	−133	−109
Ti	−41	−34	−67	−60	−36	−26	− 62	−44	− 84	− 55
V	−13	+ 2	−40	−27	+ 2	+16	− 19	+ 4	− 31	− 5
Cr	− 3	+14	−29	−16	+14	+30	− 3	+22	− 9	+ 15
Mn	−20	− 8	−43	−34	−11	+ 2	− 27	− 7	− 35	− 14
Fe	− 5	+14	−32	−18	+12	+29	− 2	+24	− 4	+ 18
Co	−18	− 2	−42	−31	− 5	+11	− 16	+ 8	− 16	+ 2
Ni	−23	− 8	−47	−37	−12	+ 4	− 22	+ 2	− 20	− 4

[10] In thermodynamic equations we are always concerned with $\Delta G \equiv \Delta H - T\Delta S$, i.e. with the change in the free enthalpy G rather than with the change in H. In the formation of alloys of two metals, however, there is little difference between ΔH and ΔG: the entropy effect ΔS is relatively small. In the case of oxides ΔS does play a part because of the considerable entropy of oxygen in the gas phase. Thermodynamic data for metal oxides have been well established from experiment.

[11] This reaction has not been verified experimentally.

[12] The value of $\Delta H(GdNi_5)$ can be found by reading off the value for Y and Ni in Table I and finding the value of the 1:5 concentration from fig. 5 or fig. 6. Because of the volume contractions in the series of rare-earth metals, Gd (in the middle of the series) resembles Y more than it does La, whereas Lu shows more resemblance to Sc; the first half of the group of rare-earth elements lies between Y and La.

[13] L. Brewer and P. R. Wengert, Met. Trans. 4, 83, 1973.

[14] K. A. Gschneidner, Jr., J. less-common Met. 17, 1, 1969.

[15] A. R. Miedema, J. less-common Met. 46, 167, 1976.

Doubtful cases, in which the two modifications of Eu or Yb have nearly the same energy, are important because exceptionally strong magnetoelastic couplings or high thermoelectric voltages may then occur. From the ΔH values given in Table I it is possible to predict in which metal systems and at what concentrations of Yb or Eu these borderline cases may be found. The amount of detail that these predictions can supply appears from the following example of Yb-Ni alloys [16]; see *fig. 9* and *fig. 10*.

In fig. 9 curve *2* gives the predicted values of ΔH for compounds in the Yb-Ni system for Ca-like Yb; curve *3* gives the concentration dependence of ΔH for Yb as a normal rare-earth metal. The trivalent compounds are more stable; a correction still has to be made, however, for the heat of conversion (33 kJ) from divalent to trivalent Yb. With this correction included, curve *3* changes to 3^{corr}; the difference between *3* and 3^{corr} is $33c_{Yb}$. We see now that on the nickel-rich side the trivalent modification remains the more stable one, but that at a concentration of more than 80% Yb the divalent modification has the lower enthalpy. Nevertheless the prediction is that there will be no divalent Yb compounds. When an alloy of say 80% Yb is dissociated into one with 60% Yb and the same quantity of pure Yb, the enthalpy is lowered (the dashed line in fig. 9). This is exactly what happens experimentally, as can be seen from the phase diagram of the Yb-Ni system (fig. 10).

Hydrogen-absorbing intermetallic compounds

Some intermetallic compounds of transition metals have the property of being able to absorb large quantities of hydrogen from the gas phase. This effect in $LaNi_5$ has been described in a previous article in this journal. The reaction

$$LaNi_5 + 3\ H_2(gas) \rightarrow LaNi_5H_6 \qquad (5)$$

is reversible at room temperature at an equilibrium pressure for hydrogen gas in the region of 10^5 Pa (i.e. 1 atmosphere). As can be seen in *fig. 11*, the six hydrogen atoms per $LaNi_5$ molecule are all absorbed at virtually the same pressure, which is a very important feature in practice. The pressure increases rapidly with increasing temperature [18].

The hydrogen-absorption effect has been studied in detail for compounds of the $LaNi_5$ family [19]. The positions of the hydrogen atoms in the $LaNi_5$ crystal lattice are known, and the behaviour when other metals are completely or partially substituted for Ni or La is known. Our knowledge of the heats of formation of alloys now enables us to begin to explain these effects.

We proceed from three initial assumptions. The first is that it is broadly sufficient if the ΔH can be predicted for reaction equations of type (5). An equilibrium pressure of 1 atmosphere means by definition that for this reaction the change ΔG in the free enthalpy is equal to zero:

$$\Delta G = \Delta H - T\Delta S = 0. \qquad (6)$$

Now the entropy effect ΔS for hydride formation at

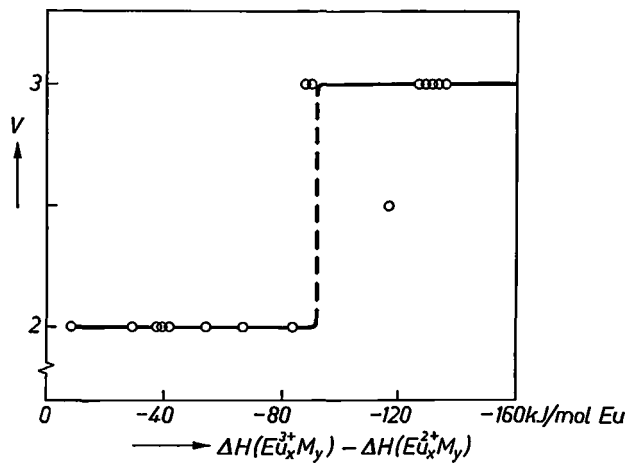


Fig. 8. The relation between the number of conduction electrons V for Eu atoms in intermetallic compounds and the calculated difference in heat of formation for compounds of the divalent (Ca-type) and trivalent (Y-type) form. It can be seen that there is a critical value for the difference in ΔH (about −90 kJ/g.at Eu), which separates the divalent from the trivalent Eu compounds.

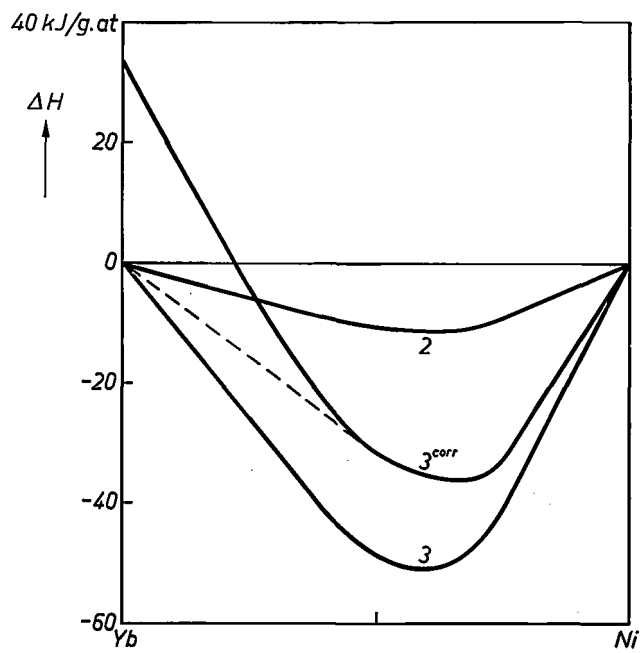


Fig. 9. Comparison of the stability of divalent and trivalent intermetallic compounds of Yb in the system Yb-Ni. The figure shows ΔH (in kJ/g.at alloy) for divalent Yb, for trivalent Yb and for trivalent Yb compounds, corrected with a value of 33 kJ/Yb for the heat of conversion from divalent to trivalent Yb. The prediction is that intermetallic compounds of trivalent Yb will exist at the Ni-rich side of the system; on the Yb-rich side of the system there are no stable alloys.

room temperature is more or less established. On the left-hand side of the reaction equation we have hydrogen gas with a high entropy, which is not present for hydrogen in the solid phase. This means that ΔS is approximately equal to the known entropy of hydrogen gas at room temperature (about 120 kJ/mol H_2). The condition $\Delta G = 0$ at $T = 300$ K thus corresponds to $\Delta H \approx -36$ kJ/mol H_2.

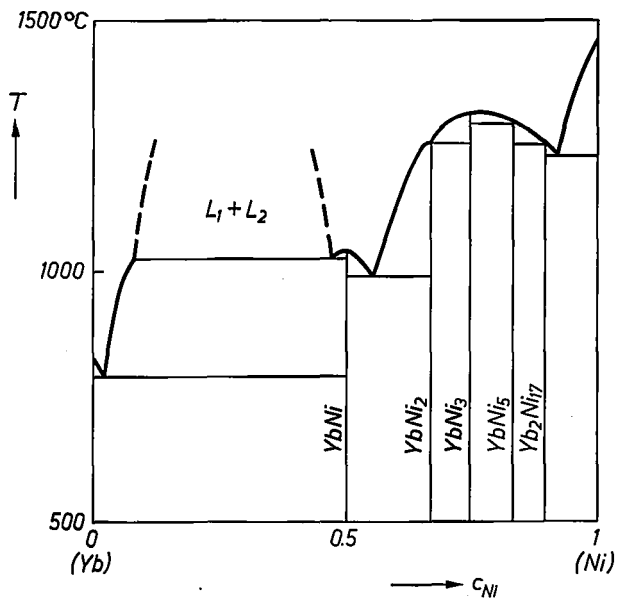


Fig. 10. Phase diagram of the system Yb-Ni, drawn from data reported by A. Palenzona and S. Cirafici [17]. There are five intermetallic compounds, in which Yb is magnetic and acquires the volume possessed by the other rare-earth metals in compounds of this type (i.e. trivalent Yb). In complete accordance with the prediction in fig. 9 phase separation occurs in the liquid on the Yb-rich side (liquid L_1 with about 8% Ni and liquid L_2 with 45% Ni at 1050 °C). Systems also exist in which Yb is always divalent (Yb-Ag) or in which Yb changes valency in mid-system (Yb-Au).

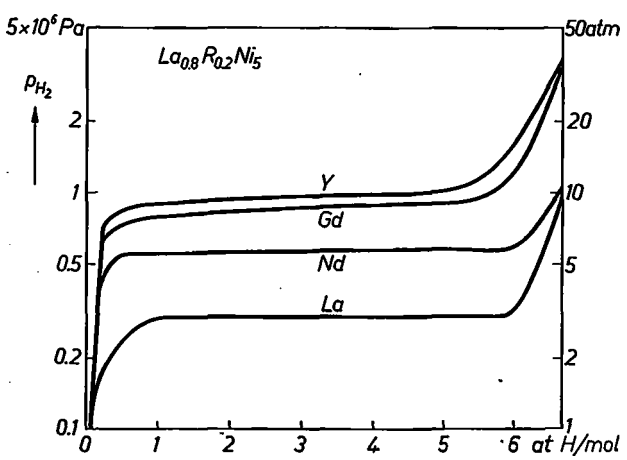


Fig. 11. The absorption of hydrogen at 40 °C by $LaNi_5$ and by analogous compounds in which 20% of the La is replaced by Y, Gd or Nd. Experimental data from H. H. van Mal [19]. For $LaNi_5$ at 40 °C the equilibrium pressure is 3×10^5 Pa (i.e. 3 atm).

The second initial assumption is that the hydrogen in the intermetallic compound will collect around the kind of atom which, as an element, forms the most stable binary hydride. This is A in *fig. 12*, and La in $LaNi_5$; La forms a very stable hydride LaH_2 or LaH_3, and nickel only just reacts with hydrogen.

Thirdly we assume that hydrogen, when it has been taken into the intermetallic compound, has itself become a metal, just as Li with one conduction electron is a metal. For hydrogen metal the rules mentioned in the second section of this article are therefore applicable: energy effects occur where atoms of dissimilar kinds are adjacent.

The diagram of fig. 12 can now be used in estimating ΔH for the ternary hydride. Fig. 12*a* shows a schematic diagram of the structure of $LaNi_5$, which owes its negative heat of formation to the contacts between A (= La) and B (= Ni) atoms [20]. In the hydride $LaNi_5H_6$, represented schematically in fig. 12*b*, the A atoms are covered by a 'blanket' of H atoms. There are now contacts between A atoms and hydrogen, and between B atoms and hydrogen, but the contacts initially present between A and B atoms have vanished. We thus arrive at the following expression for the value of ΔH for the formation of $LaNi_5H_6$ from $LaNi_5$ and H_2(gas):

$$\Delta H = \Delta H(LaH_3) + \Delta H(Ni_5H_3) - \Delta H(LaNi_5), \quad (7)$$

or more generally, for the formation of AB_nH_{2m} from AB_n and mH_2:

$$\Delta H = \Delta H(AH_m) + \Delta H(B_nH_m) - \Delta H(AB_n). \quad (8)$$

The symmetrical distribution of hydrogen over La and Ni is of course rather arbitrary, so that the result of (7) can be no more than an estimate. Nevertheless the estimate is useful, for example in describing the effect of partial substitution of La or Ni by other metals.

In applying (7) we use ΔH values for binary hydrides that have either been established experimentally or can be estimated from experiment, and also model predictions for ΔH of the starting compounds. Thus relation (7) gives: $\Delta H(LaNi_5H_6) = -250 + 0 + 67$, which is −61 kJ per mole of H_2. This is more (negative) than the expected value above of −36 kJ per mole of H_2, but it is of the right order of magnitude.

[16] This elegant example is due to Dr F. R. de Boer and Dr P. F. de Chatel of the University of Amsterdam.

[17] A. Palenzona and S. Cirafici, J. less-common Met. 33, 361, 1973.

[18] See fig. 15 in the article [1] by Van Vucht.

[19] J. H. N. van Vucht, F. A. Kuijpers and H. C. A. M. Bruning, Philips Res. Repts. 25, 133, 1970.
F. A. Kuijpers, Philips Res. Repts. Suppl. 1973, No. 2.
H. H. van Mal, Philips Res. Repts. Suppl. 1976, No. 1.

[20] The drawing agrees very well with the crystal lattice of $LaNi_5$, as shown in fig. 6 in the article [1] by Van Vucht.

If we now substitute, say, Y for La, then there is hardly any change in the first number in relation (7) ($\Delta H(YH_3) = -265$) and there is no change in the second number, whereas the third number changes markedly. Table I gives —17 kJ/g.at for LaNi and —43 kJ/g.at for YNi; taking into account a strongly asymmetric concentration-dependence as shown in fig. 6, then this amounts to —67 kJ for $LaNi_5$ and to —159 kJ for a molecule of YNi_5. Thus, YNi_5H_6 is considerably less stable than $LaNi_5H_6$. In accordance with this it can be seen in fig. 12 that the 20% substitution of Y for La increases the pressure by a factor of three. As noted earlier, Gd will have to resemble Y more closely, while Nd should lie between La and Gd. In fig. 12 this is indeed the case. H. H. van Mal and K. H. J. Buschow [21] have also performed experiments in which some of the nickel in $LaNi_5$ was replaced by Pt, Pd, Co, Fe or Cr, or some of the La by Th or Zr. In all these cases the first two terms of equation (7) undergo relatively little change, so that the result is mainly determined by the change in the third term. In Table I it can be seen that the compounds of La with Pd, La with Pt, Th with Ni and Zr with Ni are substantially more stable than those of La with Ni. The ternary hydride is therefore less stable (minus sign in equation 7) and the pressure is higher; see *fig. 13*. On the other hand, compounds of La with Co, Cr and Fe are less stable than those with Ni. (ΔH for La with Fe and Cr is in fact positive, so that starting from $LaNi_5$ we can only replace some of the Ni by Fe and Cr, and cannot make $LaFe_5$ or $LaCr_5$.) As can be seen in fig. 13, the substitution of Co or Cr for nickel does indeed give a lower equilibrium pressure, and hence a more stable hydride.

The approach used here for predicting whether or not hydrides of intermetallic compounds will be stable has also been used outside the group of $LaNi_5$-type compounds. It has been found to be a useful 'recipe', which has led to the discovery of a large number of new hydrogen-absorbing materials [22].

Surface segregation in alloys

The composition of the first atomic layer at the surface of an alloy can deviate considerably from the average composition [23]. For example, in a solid solution of Cu and Fe with a concentration of 4 at% Cu, the concentration of Cu in the first atomic layer will be considerably higher than 4%. In itself the effect is virtually self-evident; 'forming part of the surface' is obviously something like partial evaporation, since atomic cells of the first layer are only partly surrounded by other metal atoms. Consequently an atom at the surface acquires a higher energy; a reasonable assump-

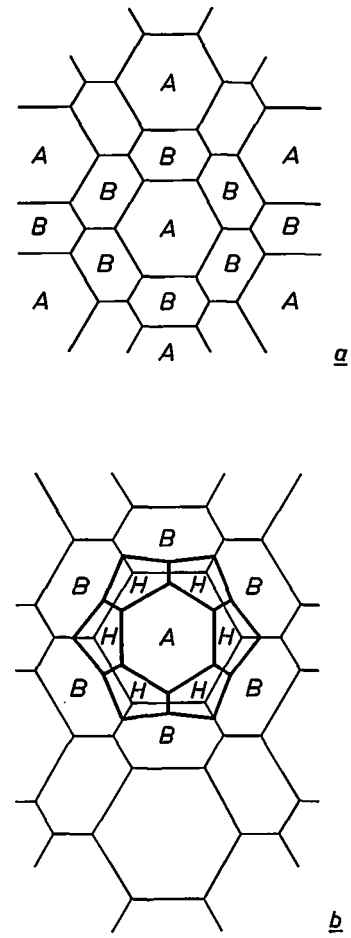


Fig. 12. Schematic representation of the formation of a ternary hydride, starting from a binary compound AB_n ($n \geqslant 1$) of transition metals. It is assumed that metal A is strongly attractive to hydrogen, and B much less so. On absorbing hydrogen the lattice expands, but diffusion of A and B atoms is not possible because of the low temperature.

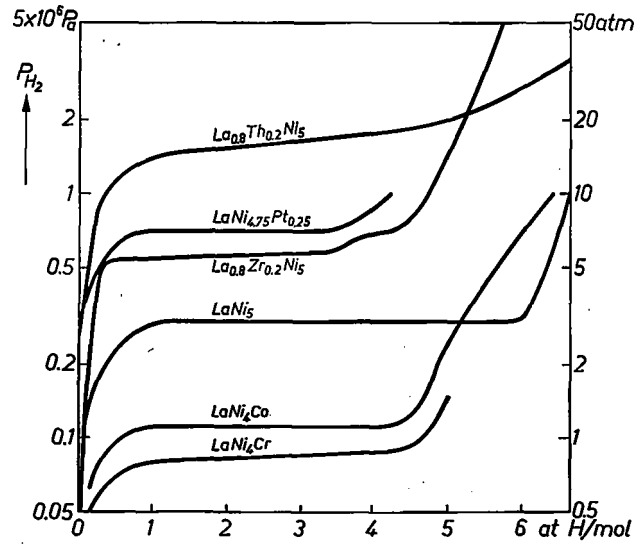


Fig. 13. The change in the equilibrium pressure for hydrogen absorption in $LaNi_5$ on partial replacement of Ni by Cr, Co or Pt, or partial replacement of La by Zr or Th.

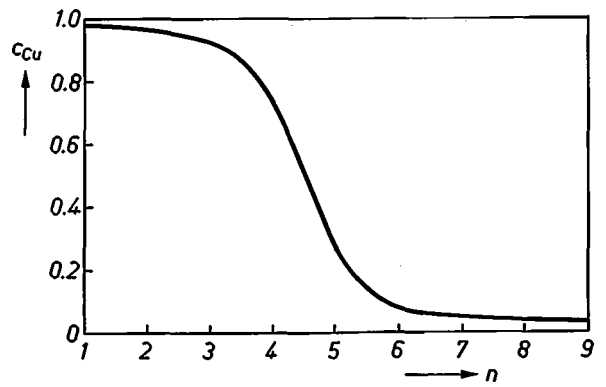


Fig. 14. Concentration profile for the copper content at the surface of an iron-rich alloy of copper and iron ($Fe_{96}Cu_4$). The calculation (due to J. L. Meijering [24]) applies in principle to a (100) surface. n refers to the numbering of the atomic layers, starting from the surface.

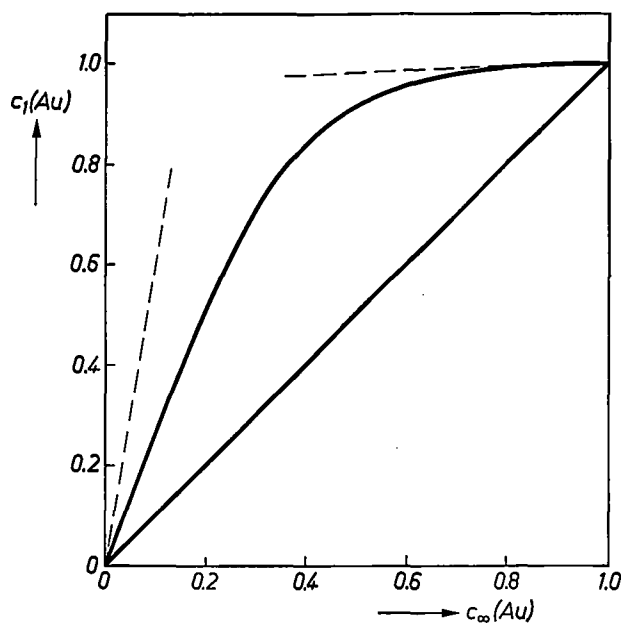


Fig. 15. The concentration of Au in the first atomic layer at the surface of Cu-Au alloys, compared with the concentration in the bulk of the material. The experimental curve was derived from Auger measurements by J. M. McDavid and S. C. Fain [25], the dashed lines were calculated from equation (10) for dilute alloys. In addition to the values for E_v^A, E_v^B and $(V_A/V_B)^{2/3}$, mentioned in the text, the calculation was made for $\Delta H_{sol} = -20$ kJ/mol and $T = 600$ K.

tion is that the energy difference is about a third of the latent heat of vaporization. In an alloy the metal with the lowest latent heat of vaporization will be the least disinclined to occupy a position at the surface, so that the concentration of this metal in the first layer could be greater than in the bulk of the material.

For our example, Cu dissolved in Fe, the concentration profile at the surface was calculated as long ago as 1966 by J. L. Meijering [24], see *fig. 14*. The composition in the bulk is Cu_4Fe_{96}. In the first layer there is much more copper ($Cu_{98}Fe_2$), and in the first five atomic layers this excess of copper rapidly decreases. For dilute alloys Meijering was able to show that

$$c_1^A/c_\infty^A = \exp[(\Delta H_{sol}^A - E_v^A + E_v^B)/3RT], \quad (9)$$

where c_1^A is the concentration of metal A in the first layer, c_∞^A is the bulk concentration, ΔH_{sol}^A is the heat of solution of A in B, while E_v^A and E_v^B are the latent heats of vaporization of the two metals. The factor 3 in the denominator indicates that an atom in the surface loses a third of its contact with neighbouring atoms, i.e. a third of its binding energy.

For a solution of Cu in Fe we have $E_v^A = 336$ kJ/mol, and $E_v^B = 415$ kJ/mol, while the heat of solution of Cu in Fe (see Table I) is positive and equal to $\Delta H_{sol}^A = 56$ kJ/mol. The signs in equation (9) are now easily understood. Since the heat of evaporation of copper is lower than that of iron, c_1^A will tend to be greater than c_∞^A. This effect is reinforced by the positive heat of solution: Cu does not want to be surrounded by Fe. The fact that Cu does dissolve in iron (up to 4%) is connected with the associated reduction in entropy. For $T = 1000$ K (i.e. $3RT = 25$ kJ) we find $c_1^A/c_\infty^A = \exp 6$, a very considerable increase in the concentration of Cu in the first layer.

If we take the converse problem of a dilute solution of Fe in Cu, we find that the two contributions oppose one another. On the one hand the minority (Fe) atoms will now avoid the surface because of the greater latent heat of vaporization, and on the other Fe at the surface will be less completely surrounded by Cu atoms, which lowers the energy of the alloy because of the positive heat of solution. We then have $c_1^{Fe}/c_\infty^{Fe} = \exp(-1.6)$. Although there is now some lowering of the Fe concentration at the surface of copper-rich/copper-iron alloys, the effect is relatively small. The marked asymmetry we encounter here emphasizes the important role played by the heat of solution in the segregation problem.

Surface segregation can be observed experimentally in various ways. An example is the composition of the first atomic layer in Cu-Au alloys, as derived from measurements of the intensity of Auger lines by J. M. McDavid and S. C. Fain [25]. *Fig. 15* shows that in the whole concentration region the Au content in the first atomic layer is considerably higher than in the deeper-lying zones.

[21] K. H. J. Buschow, H. H. van Mal and A. R. Miedema, J. less-common Met. 42, 163, 1975.
[22] A. R. Miedema, K. H. J. Buschow and H. H. van Mal, J. less-common Met. 49, 463, 1976.
[23] Prof. Dr H. H. Brongersma of our Laboratories drew our attention to this problem. Dr J. J. van den Broek showed how the heat of formation (in this case the heat of solution) should be used here.
[24] J. L. Meijering, Acta met. 14, 251, 1966.
[25] J. M. McDavid and S. C. Fain, Jr., Surface Sci. 52, 161, 1975.

The remarkable thing about this result is that it does not at first sight seem to tally with the values of the latent heats of vaporization: that of Au (367 kJ/mol) is higher than that of Cu (336 kJ/mol). The explanation lies in the difference in the size of Cu and Au atoms,

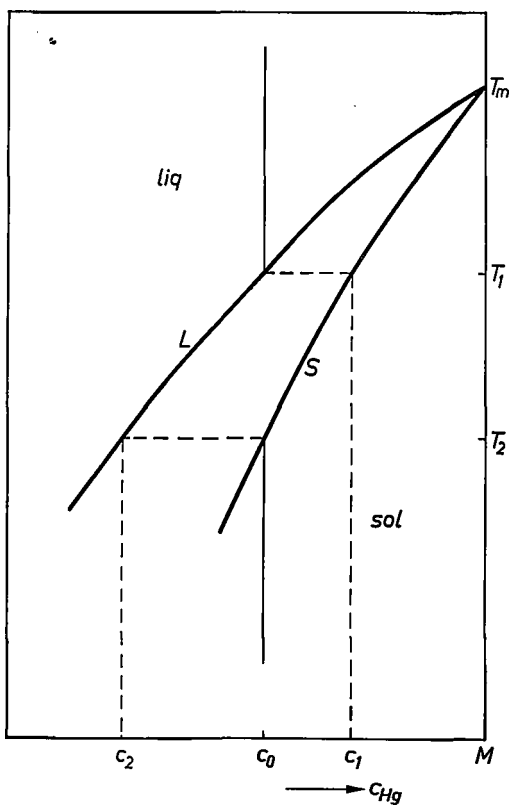


Fig. 16. The solidification curve of a dilute solution of Hg in a metal M. An alloy of M (melting point T_m) with Hg has a lower solidification point than pure M. When, upon decreasing temperature, a liquid of composition c_0 reaches the temperature T_1, a solid crystallizes out in which the Hg concentration is equal to c_1 ($c_1 < c_0$). On further decrease of temperature the composition of the liquid consequently changes in the direction of more Hg, until at T_2 all the liquid has solidified. The vapour pressure of Hg above the alloy is given by:

$$p_{Hg} = c_{Hg} \exp - (\Delta H_v + \Delta H_{sol})/RT.$$

Between T_1 and T_2 the first term increases and the second term decreases, making it possible to obtain a maximum in p_{Hg} as a function of temperature. The presence of ΔH_{sol} in the exponent allows the value of p_{Hg} [27] to be regulated.

which amounts to a factor of 1.44 in volume (in the previous examples Cu and Fe were of the same size). To cover a given surface with Cu, more atoms are needed than to cover it with Au. Equation (9) must therefore be changed to:

$$c_1^A/c_\infty^A = \exp\,[\{\Delta H_{sol}^A - E_v^A(V_B/V_A)^{\frac{2}{3}} + E_v^B\}/3RT], \quad (10)$$

where V_A and V_B are the molar volumes of the two metals. For Cu in Au we find that $(V_B/V_A)^{\frac{2}{3}}$ is equal to 1.28, a factor which means that it is now relatively cheaper to have Au atoms at the surface, in agreement with the experiment.

The concentration profile and the associated variation in properties can have an effect on the corrosion and adhesion properties of alloys, the value of their work function, the energy for vacancy formation and the energy of grain boundaries.

Other examples

Predictions of the heats of mixing of liquid metals can also be of great practical value. An example is to be found in the control of the vapour pressure above liquid metals, in particular that of mercury in mercury alloys. For applications in fluorescent lamps it is most desirable to have a mercury alloy whose vapour pressure is practically constant over a temperature range of a few tens of degrees. In normal cases the vapour pressure of a substance rises very rapidly with increasing temperature (the curve is exponential). In a special case, that of a liquid dilute mercury alloy moving along the line of fusion (see *fig. 16*) because of equilibrium with the associated solid phase, it has been possible to give the location of a maximum in the vapour pressure as a function of T [26] [27]. Since, apart from the presence of a maximum in the vapour pressure, the specification also stipulates the value of the vapour pressure at this maximum (0.8 Pa, i.e. 6×10^{-3} torr), it is important to know the heat of solution of Hg in addition to the latent heats of fusion and the melting points of the metals (Hg and the solvent) to enable the specifications to be met.

Another technical problem where a knowledge of the heats of mixing of liquid metals demonstrates its value is that of cooling circuits (nuclear reactors, Stir-

Table III. Heat of solution ΔH_{sol} of a number of solid transition metals A in low-melting liquid metals M_{sol}. In the absence of compounds in the system the solubility of a metal A is given to a good approximation by: $c_A = \exp(-\Delta H_{sol}/RT)$. The unit of ΔH_{sol} is kJ/mol, the value is always positive here (R = 8.3 J/mol).

A \ M_{sol}	Na	Mg	In	Tl	Sn	Pb	Bi
Cr	235	96	83	117	46	101	88
Mn	169	50	25	55		38	25
Fe	211	78	78	115	48	107	95
Co	141	24	38	70	16	70	58
Ni	117	5	21	59	4	59	46
Mo	342	165	144	186	103	170	153
W	361	181	169	214	127	202	185

ling engines) in which the coolant is a liquid metal. The coolant most commonly used is sodium. If, however, the use of other low-melting metals is contemplated (In, Tl, Sn, Pb, Bi), the choice can be facilitated by a knowledge of the heats of solution of transition metals such as Cr and Fe in these liquids. A survey of these data is presented in *Table III*.

[26] G. Franck, Z. Naturf. 26a, 150, 1971.
[27] J. Bloem, A. Bouwknegt and G. A. Wesselink, to be published in J. Illum. Engng. Soc.

Summary. A model can be used to predict the heat of formation for any binary alloy that contains at least one transition metal (both liquid alloys and solid intermetallic compounds). As in chemistry generally, the heat of formation is an important quantity for metals. In this article it is demonstrated that the predictions, which link up excellently with all the available experimental data on heats of formation and phase diagrams, do not always correspond to expectations based on the regularities of the periodic system. Some examples of metal problems in which it is vital to know the heats of formation of alloys are discussed in detail. These include the changing valency of the rare-earth elements Eu and Yb in the metallic state, the reversible absorption of hydrogen in intermetallic compounds, and the segregation that can occur in the first atomic layers at the surface of an alloy.





The video telephone can be used for the discussion of a production schedule with outlying factories and suppliers.

Digital circuits in the video telephone

H. P. J. Boudewijns, E. C. Dijkmans, P. W. Millenaar, N. A. M. Verhoeckx and C. H. J. Vos

A previous article [1] gave a general description of the experimental video-telephone system that has been in use at Philips Research Laboratories in Eindhoven since 1972 and is also connected to several other places in the Netherlands. An interesting feature of the signals used in this system is that they alternate from analog to digital; the video signals are analog, whereas the synchronization, speech-transmission and signalling signals are digital. This alternation in signal type is rather unusual, and this is why we are including the special article on the function of the digital circuits in the video telephone.

Features of a video-telephone system

In a video-telephone system three kinds of signal have to be transmitted. First there is the video signal. This requires by far the largest bandwidth, which is about 1 MHz in the Dutch experimental system [1], and in other systems. At each terminal a video signal is generated simultaneously for transmission to the other terminal. There is also the speech signal, occupying at least telephony bandwidth, and thirdly there are signals for communicating with the exchange (dialling signal, calling signal and various tones); the whole of this last group is used for the 'signalling'.

The last two kinds of signal are also found in ordinary telephone practice. It seems fairly obvious that they can be transmitted over a standard telephone circuit, and in fact this is what is done in some video-telephone systems. The video circuit is then an addition to the existing telephone circuit. The separate exchange necessary for switching the wideband circuit is controlled from the existing telephone exchange. The subscriber wishing to make a video-telephone call can for example indicate this by dialling certain digits before the ordinary telephone number. This system has the advantage that if the video circuit develops a fault, the speech circuit and signalling are not affected. A power supply failure at either end of the video line is sufficient to make the picture disappear, since each video telephone is powered from its local mains. An ordinary telephone, on the other hand, derives its current from the exchange, where there is always emergency equipment to ensure that it continues to function if the public power supply fails.

The introduction of a video-telephone system with a structure of this kind grafted on to the public telephone system depends upon the existence of modern telephone exchanges controlled by programmed processors. The number of such exchanges is growing but it will be many years before they are general throughout the system.

Another aspect is the quality of the sound. The video telephone requires a loudspeaking telephone circuit. The quality of the sound is considerably improved if it is given a bandwidth larger than the 3.5 kHz normally used in telephony.

A way of doing this has been devised for use in the experimental system in the Netherlands. It consists in coding the sound in digital form and transmitting it during the line-blanking intervals of the video signal. This has a second important advantage: the sound signal is regenerated at the receiver and the level of the sound does not therefore depend on the line attenuation. In particular, this simplifies the voice-actuated switch in the loudspeaking telephone circuit.

In the cross-country sections of the experimental system a telephone line for signalling runs alongside the video pairs. At Philips Research Laboratories, however, where a special video-telephone exchange had to be built, it was decided to incorporate the signalling in the video signal in digital form, thus enabling the extra telephone line to be eliminated. This means that the entire video-telephone circuit can be accommodated in the four conductors of the outgoing and incoming video pairs.

H. P. J. Boudewijns, formerly with Philips Research Laboratories, is now with the Philips Commercial Estimating and Planning Department, Eindhoven; E. C. Dijkmans, P. W. Millenaar and Ir N. A. M. Verhoeckx are with Philips Research Laboratories; Ir C. H. J. Vos, formerly with Philips Research Laboratories, is now with the Philips Audio Division, Eindhoven.

[1] E. A. Aagaard, P. M. van den Avoort and F. W. de Vrijer, An experimental video-telephone system, Philips tech. Rev. 36, 85-92, 1976 (No. 4).

The use of digital signals creates a number of new possibilities. With digital transmission it is of course necessary to use synchronous bit-clock oscillators at the transmitting and receiving stations, and these oscillators can obviously also be used for field and line synchronization. It has been found possible to develop a system for this that is extremely insensitive to noise and interference. Moreover, the field-blanking period is long enough to permit the transmission of additional digital information, a facility that can be put to various good uses.

When a subscriber dials a number the exchange has to decode the digital dialling information it receives. Previous synchronization is essential here, and fairly complicated receivers are therefore required at the incoming side of the exchange. To limit expense each subscriber can be supplied with an 'off-hook detector' instead of a complete receiver, and receivers can be allocated to calling subscribers as required.

This article presents a description of the digital aids used in video-telephone sets in the experimental system. It will be concerned chiefly with two aspects: the synchronization of the receiver in one video-telephone set with the transmitter in the other set, and digital speech transmission with the voice switch. Finally, some extensions of the application of the system that are largely due to the digital approach will be considered.

Block diagram of the video-telephone set

A general description of the use of digital signals in the video-telephone set will now be given with the aid of the block diagram in *fig. 1*, where the circuits operating with analog signals are shown in black and the digital circuits in colour (red for outgoing, blue for incoming signals). It will be observed that only the video signal from the camera *Cam* retains its analog form, and this accounts for the analog components in the outgoing line signal. All other signals are given a digital form.

The time scale of all signals proceeding in a particular direction is determined by the oscillator *G*, which generates the clock signal in the transmitter. The clock signal is taken to a chain of frequency dividers *DivCh;* the results of these divisions are combined in logic circuits *TxLog* to give the 21 control signals required for the various operations in the transmitter.

These include the picture scanning in the camera. The clock rate (1.0016 MHz) has been chosen such that the line frequency (7825 Hz) and the picture rate (25 Hz) can be derived from it by division. Another function performed by the control signals is the sampling of the speech signal; this is done eight times per picture line in the delta modulator *DMod*, and depending on the sign of the difference between successive samples a '1' or a '0' is generated. The bit rate is 62.6 kbit/s, which is adequate for a sound bandwidth of about 5 kHz; this value corresponds to the bandwidth used for radio broadcasts in the medium-wave band. The eight bits generated per picture line are accumulated in a speech buffer *SpB* and transmitted at an accelerated rate in the blanking interval at the end of the picture line.

The information signalled by the subscriber when he uses pushbuttons *PBut* is coded digitally; it is then stored in a signalling buffer *SigB* and transmitted to the exchange in the field-blanking intervals.

A third digital signal, of great importance for synchronization between transmitter and receiver, is generated in the synchronization-word generator *SyWG*. This is a 32-bit word that is transmitted once per field; the word in the odd field is the logical inverse of the word in the even field. The three digital signals are combined in a digital multiplexer *DMux* with trains of square-wave pulses at half the clock rate in the line- and field-blanking intervals; these pulses are used for bit-clock synchronization. The output signal is again combined with the analog video signals in an analog-digital multiplexer *ADMux* and the combined signal is then transmitted.

At the same time a signal of similar composition is received over the incoming line. The analog components of this signal modulate the beam current in the picture tube. First, the digital components are used to synchronize the receiver. The synchronization signal is produced by preamplifying the entire signal, 'clamping' it to restore the d.c. voltage level, and then clipping it severely. The video information is lost in this process. The resultant signal is applied to various stages, each of which processes part of the information it contains.

Two of the stages are for synchronization. The bit-clock oscillator of the receiver reacts to the trains of square-wave pulses in the incoming signal. Their phase is compared with that of the locally generated clock pulses and from the comparison a d.c. voltage arises that adjusts the frequency and phase of the oscillator (phase-locked loop, block *PLL* in the figure). This gives synchronization at the highest frequency (1.0016 MHz). There is also a synchronization-word detector *SyWD* that emits a pulse on reception of a synchronization word. If this pulse exceeds a certain threshold in the synchronization verification stage *SyVer*, it resets the counters in the divider chain *DivCh*. The synchronization words thus act as a time marker recurring at the field frequency (50 Hz). (In fact the synchronization words for odd and even fields are different, thus uniquely specifying the picture frequency of 25 Hz.) All intervening frequencies, generated in the divider chain

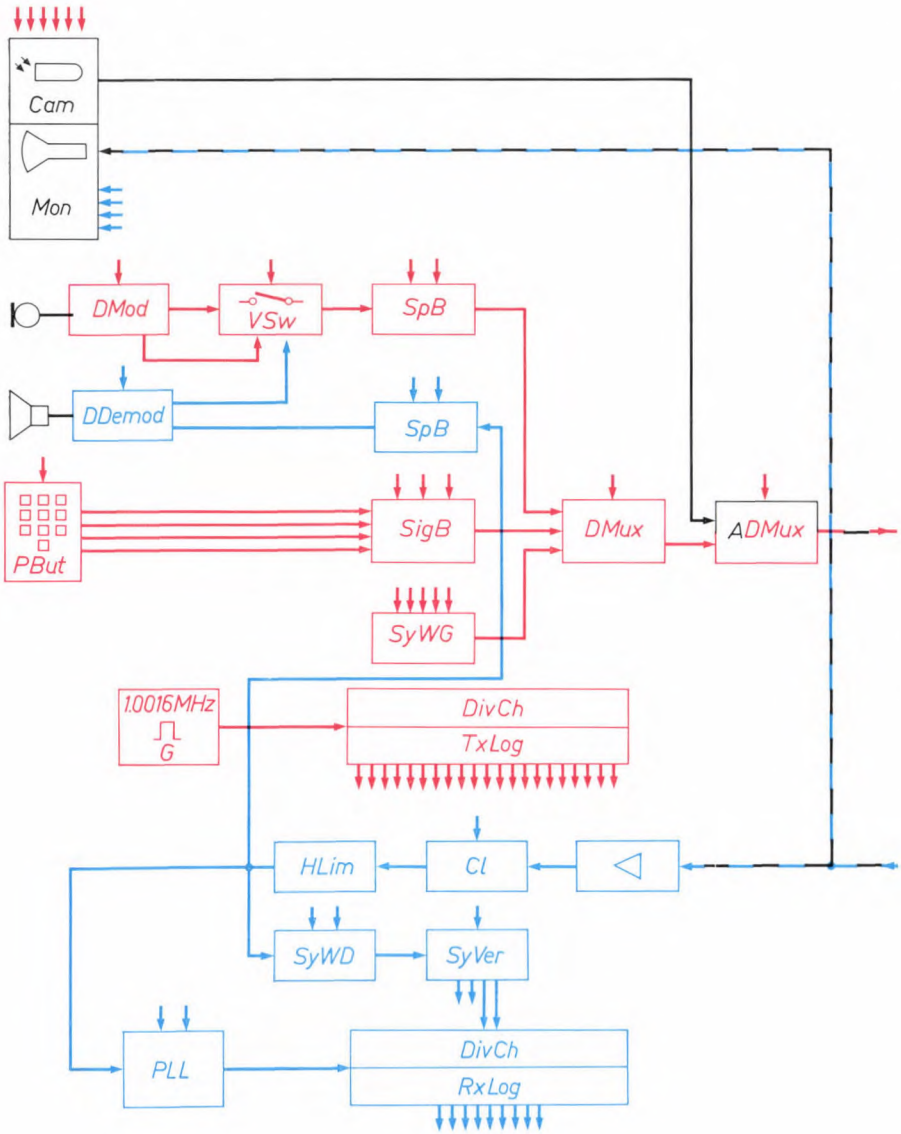


Fig. 1. Block diagram of the video-telephone set. *Black:* analog signals. *Red:* digital signals from the transmitter. *Blue:* digital signals intended for the receiver.

Cam	camera
DMod	delta modulator
VSw	voice-actuated switch
SpB	speech buffer
PBut	pushbuttons for subscriber dialling
SigB	signalling buffer
DMux	digital multiplexer
ADMux	analog-digital multiplexer
SyWG	synchronization-word generator
G	bit-clock generator
DivCh	divider chain
TxLog	transmitter logic
Mon	monitor
DDemod	delta demodulator
Cl	clamp circuit
HLim	voltage limiter
PLL	bit-clock oscillator, adjusted by phase comparison
SyWD	synchronization-word detector
SyVer	synchronization verification
RxLog	receiver logic

of the receiver, are then available in the correct phase and the receiver logic *RxLog* combines them into the required control signals.

These are used for decoding the speech signals. The incoming eight-bit words are stored in a buffer *SpB* and taken in the form of a continuous stream of bits to the delta demodulator *DDemod*, at whose output the analog sound signal becomes available. The delta demodulator also provides a signal that indicates the amplitude of the incoming speech signal and is compared in the voice-actuated switch *VSw* with a corresponding signal from the delta modulator *DMod*. If the signal from *DMod* is the weaker of the two, the microphone channel is disconnected. This ensures that there is no closed electro-acoustic circuit in which acoustic feedback due to the built-in amplification might lead to oscillation.

The receivers forming part of the exchange equipment are basically the same as those in the video-telephone set. However, no optical and acoustical output stages are included and they incorporate some additional units for processing the signalling information.

The block diagram of the video-telephone set in fig. 1 has been simplified by the omission of two subassemblies: the 'self-view' switch, which enables the local subscriber to monitor the picture being transmitted, and a test-signal switch, which generates a checkerboard pattern from the existing signals. This pattern is also transmitted if the subscriber presses a special button that prevents the transmission of his picture.

Synchronization

Synchronization of the bit-clock oscillator

As we noted earlier, the receiver is synchronized to the incoming signal at both ends of the frequency spectrum: at the clock rate (1.0016 MHz) and at the picture

frequency (25 Hz). The bit-clock oscillator is synchronized by using series of alternating 'zeros' and 'ones', inserted into the signal at fixed positions by the transmitter. These form a square wave at a frequency of 500.8 kHz. Six cycles, i.e. twelve bits, are accommodated in every line-blanking interval; in the field-blanking interval five complete lines are occupied by alternating '0's and '1's (*fig. 2*).

The receiver has two synchronization modes. One of these is the acquisition mode; the receiver is in this mode when the connection is first established, and returns to it when noise or interference causes synchronization to be lost. In the acquisition mode the incoming signal is applied continuously to the bit-clock oscillator of the receiver. During this time the synchronization-verification circuit keeps a continuous watch for the detection pulse that will occur when a synchronization word is received; to prevent errors a high threshold is applied.

The other mode is the maintenance mode, to which the receiver switches when synchronization has been achieved. The bit-clock oscillator no longer has the entire signal applied to it but only the 500.8 kHz square-wave signals at the end of every line. The watch for the detection pulse is confined to the bit interval in which the detection pulse should appear; the detection threshold is lowered.

The long series of '0's and '1's during the field blanking is necessary for rapid acquisition of synchronization. This is because the received signal has to undergo a number of processing operations, as we saw earlier. The video signal has a voltage swing that is ten times that of the digital signal 'fragments' and if by chance the video signal contained a run that looked like a succession of '0's and '1's, this would certainly destroy synchronisation. To eliminate the video signal as much as possible, only a small voltage swing is permitted, as we said earlier; this is restricted to a few per cent of that of the digital fragments (hard limiting). The trains of digital pulses have previously been raised to the correct d.c. level (clamping), which is equivalent to restoring the d.c. component lost in the transmission.

The reconstitution of the d.c. level consists in determining the mean voltage of the digital-signal fragments and raising this to the nominal zero level; this also has the effect of restoring the black level of the video signal. The circuit determining the mean voltage is activated by a pulse each time the 500.8-kHz square wave is received in the line-blanking interval. If, however, the receiver has not yet been synchronized at that moment the pulse occurs at a random point in the picture line and an incorrect d.c. voltage level results. It is only during the first line period of the field blanking that the 500.8-kHz square-wave signal is received during the pulse in every case. After this has happened, the correct d.c. voltage level is set within about two line periods. An undistorted continuous 500.8-kHz square-wave signal then reaches the phase-locked loop; the oscillator is now synchronized with the incoming signal, again within about two line periods. As a general rule

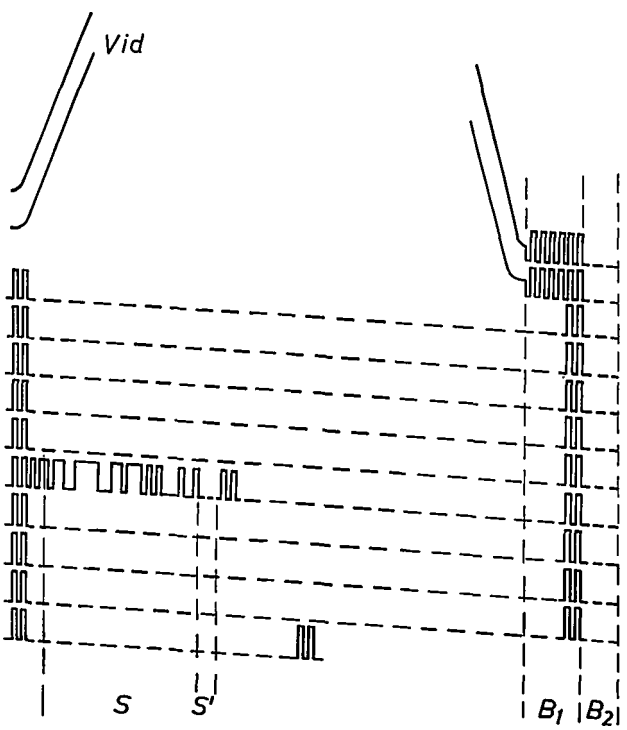


Fig. 2. The signals in the blanking interval of an even field. B_1 twelve bits for bit-clock synchronization. B_2 eight bits for sound transmission. S 32-bit synchronization word. S' four bits for dialling information. The long series '0101 . . . 01' in the first five lines are for fast acquisition of synchronization at the beginning of a call. The pulses after S' can without difficulty be replaced by information-carrying signals. *Vid* video signal with a voltage swing up to ten times that of the pulses.

the correct d.c. voltage is thus restored and the bit-clock oscillator synchronized with the received clock signal within five line periods after the start of the first field blanking.

The conditions for detection of the synchronization word are now satisfied and this word follows in the sixth line period.

Synchronization word, matched filter

Once the clock oscillator of the receiver has been synchronized, the numerous signals derived from it by frequency division are at the correct frequency, but their phase relative to the corresponding signals at the transmitter is still indeterminate. As we saw, the function of the synchronization word is to establish the correct phase relations.

These words consist of a digital signal that can be recognized clearly by the receiver even when there is noise or interference in the transmission. They have been given the form of a special sequence of '0's and '1's which is detected in the receiver with an appropriate matched filter.

cross-correlation function of the pattern built into the matched filter with the received signal — the synchronization word in a context of '0's and '1's — has a peak standing out clearly from its surroundings.

This requirement for unambiguous detection is more easily met the longer the synchronization word. There

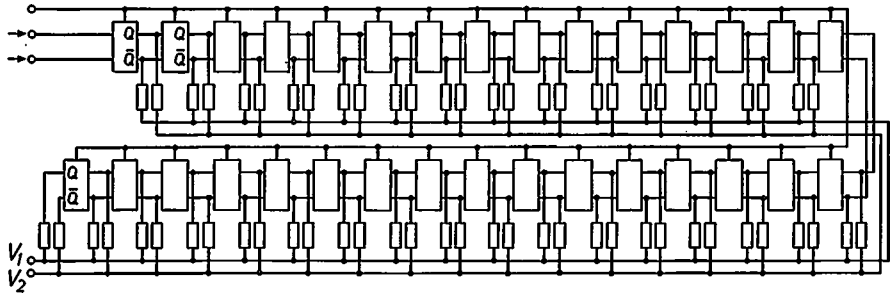


Fig. 3. The matched filter (shown simplified). This is a shift register; all the outputs (Q and $\bar{Q}$) of the succession of flipflops that carry a 'high' voltage when an even synchronization word is stored in the shift register are connected via resistors to output V_1. The other outputs then carry a low voltage and are connected to output V_2. The received signal is shifted through the matched filter. The voltage at output V_1 shows a positive peak when the even synchronization word passes through the filter, and the voltage at output V_2 a positive peak when the odd synchronization word (the logical inverse of the even word) passes through the filter.

This matched filter is a shift register with 31 stages, each of which has two outputs, Q and $\bar{Q}$; see *fig. 3*. If a stage is in state '1', a high voltage occurs at output Q and a low voltage at output $\bar{Q}$; for a stage in state '0' the situation is reversed. All the outputs are distributed via resistors in such a way over two common lines that when the shift register contains exactly one of the synchronization words all the outputs at which there is a high voltage are connected together. A maximum positive voltage then becomes available on one of the common lines, e.g. V_1. The other outputs then all carry the low voltage and an equally large but negative voltage occurs on V_2. With the other synchronization word, the polarities are exactly the opposite because this word is the logical inverse of the first. If the synchronization word in the matched filter is shifted one or more places, the voltages are smaller. The maximum voltage corresponds to the maximum correlation between the received signal and the bit pattern designed into the matched filter.

The signal entering the receiver is stepped through the matched filter after voltage limiting. If the output voltage from the filter exceeds a certain threshold, a detection pulse is generated. For optimum synchronization it is essential that this detection pulse should occur only when the matched filter contains the complete synchronization word. This word has therefore been chosen such that the output voltage from the matched filter is very much smaller at every other moment; the are, however, practical limits; in the form of signal chosen the synchronization word has to fit into half a line period (see fig. 2), part of which has already been reserved for the twenty bits required for clock synchronization and speech. Furthermore, the size of the synchronization word generator and the matched filter increases with the length of the word. We chose the 32-bit length mentioned above, i.e. a quarter of a line period.

The synchronization word consists of a single cycle of a 'maximum-length sequence' with a length of 31 bits, plus the first bit repeated. A characteristic of maximum-length sequences is that when they are periodically repeated their autocorrelation function has a definite peak in each cycle [2]. The choice of a maximum-length sequence means that of the approximately one thousand million possible 31-bit words only 186 need be considered. These are found to be cyclical shifts of six basic forms, three of which are the time inverses of the other three.

The most suitable of these 186 series was chosen with the aid of a computer program, with an 'environment' of alternating '0's and '1's as a precondition. The most suitable word was found to be:

'10110011111000110111010100001001'.

This word is used as the synchronization word in the even field and its logical inverse in the odd fields.

[2] S. W. Golomb, Digital communications with space applications, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1964.

The cross-correlation functions of these two words in an environment of alternating '0's and '1's with the bit pattern of the matched filter are shown in *fig. 4*. This figure also shows the threshold levels that the peak must exceed to cause generation of the detection pulse. This pulse resets the counters in the frequency divider chain (*DivCh*, fig. 1) and thus ensures that the phase of the signals resulting from the divisions is correct.

Verification

Verification consists in examining every detection pulse to see whether it occurs at the moment at which it can be expected on the basis of the preceding pulses. If it does, the reading of an up-and-down counter is increased by one; the maximum reading to which the counter can advance is 31. If a detection pulse fails to occur when expected, the reading of the counter is reduced by one.

The up-and-down counter determines whether the receiver is set for acquisition or maintenance of synchronization. If the counter is in any position between 16 and 31, the receiver is set for maintenance of synchronization; the signal is supplied to the phase-locked loop for only twelve bits during the line-blanking interval during which the low threshold level applies to the detection pulses. If because there are no detection pulses the counter comes to occupy any of the positions 1 to 15, the receiver has to acquire synchronization again, but the connection is maintained; the received signal is then continuously supplied to the phase-locked loop, the detection threshold is raised and the speech signals are suppressed. Only after 15 failures to restore synchronization does the counter occupy the position 0, which breaks the connection.

Speech transmission

Digitally controlled delta modulation

The choice of delta modulation (*fig. 5*) for the digital coding of the speech was determined by the comparative insensitivity of this modulation system to noise and interference. It is better in this respect than the pulse-code modulation much used in telephony, in which the magnitude of successive signal samples is expressed in digital words consisting of several bits. In pulse-code modulation the first bit of each word has a weight of half the maximum signal value, so that an error in that bit causes considerable disturbance. In delta modulation, on the other hand, an incorrect bit can never produce a deviation exceeding a single quantization unit [3].

The difference between the actual signal and the 'echelon curve' obtained with delta modulation is expressed in the effect known as quantization noise. As fig. 5 shows, the strength of this noise is independent of the signal strength; the ratio therefore does not favour weak signals. The situation can be improved by adapting the size of the quantization steps to the signal strength. This is done by compressing the signal-strength differences at the transmitter and expanding them at the receiver — the combined arrangement is known as 'companding'.

In the video telephone digitally controlled companding [4] is introduced by means of the circuit shown in *fig. 6*. In addition to the normal feedback loop to the pulse modulator *PM*, the delta modulator on the left of the figure has a second loop in which a modulation-level analyser *MLA* and a pulse-amplitude modulator *PAM* are inserted. The first loop operates as follows. The pulse modulator *PM* generates a '1' or a '0', depending on whether the echelon curve (fig. 5) has to go up or down one unit; this information is derived from the difference signal e originating in the difference amplifier D from a comparison of the generated echelon curve with the input signal v_1. The echelon curve is

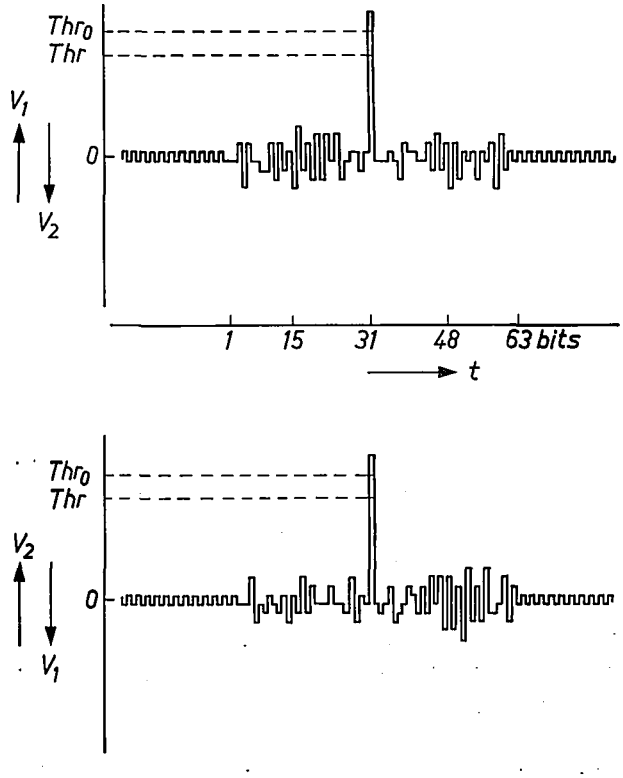


Fig. 4. The voltages at outputs V_1 and V_2 of the matched filter while the synchronization word is being received. It is assumed that the word is surrounded entirely by the '0101 . . . 01' series. *Top:* even-field blanking. *Bottom:* odd-field blanking. t time in bit units after the start of the synchronization word. *Thr* threshold which the peak must exceed before the synchronization word is accepted. Thr_0 raised threshold in force during acquisiton.

provided by an integrator for this comparison. The same integration takes place in the demodulator, where it gives the output signal.

The pulse-amplitude modulator *PAM* can now affect the size of the steps, however. It does not do so as long as the '1's are not being transmitted in groups exceeding three in succession. Then every three '1's are followed by at least one '0', which means that in four time units the maximum increase is equal to two steps. The modulation index therefore does not exceed $\frac{1}{2}$. However, as soon as more than three '1's arrive in succession, the modulation-level analyser produces a pulse. After smoothing, these pulses result in a control voltage v_c, whose effect is to make *PAM* increase the size of the steps. The tendency to form groups of more than three '1's will therefore in general decrease again and the system will thus try to maintain the modulation-index value near to $\frac{1}{2}$ over a large range of input levels; in other words there is compression. The same applies if, conversely, '0's are transmitted in groups of more than three at a time. The demodulator also includes a pulse-amplitude modulator; the step size here will accurately follow that at the transmitter since both are driven by the same signal — transmission errors excepted; the result is therefore expansion. When v_c is zero, a minimum step size is maintained at both ends because a small d.c. voltage V_0 ($\approx 0.01\ v_{c,max}$) is injected.

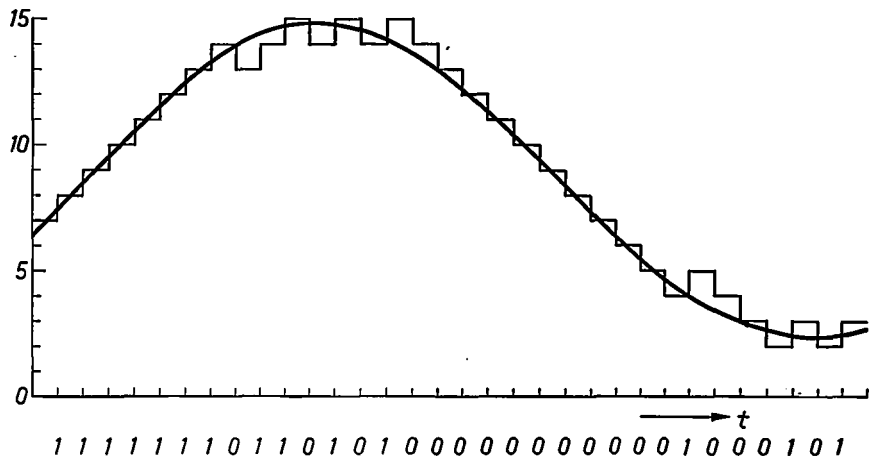


Fig. 5. Delta modulation. A '1' is transmitted when the 'echelon curve' by which the signal is approximated has to rise one step, a '0' when it has to fall one step.

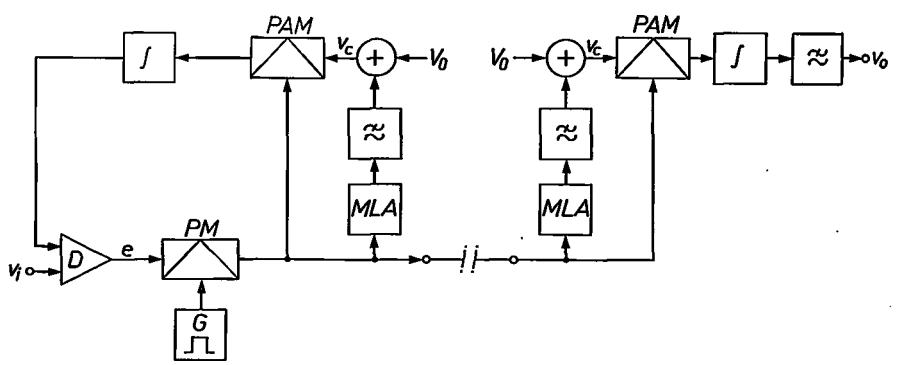


Fig. 6. Simplified block diagram of a delta modulator (*left*) and demodulator (*right*) with digitally controlled companding. This is achieved with the aid of the extra control loop incorporating modulation-level analyser *MLA* and pulse-amplitude modulator *PAM*. *D* differential amplifier. *G* pulse generator. *PM* pulse modulator. v_i input voltage. *e* difference signal. v_c control voltage. V_0 d.c. voltage for maintaining a minimum step size. v_o output voltage.

The improvement in the signal-to-noise ratio resulting from the companding can be seen from the difference between the two curves in *fig. 7*, where curve *A* represents the signal-to-noise ratio without and curve *B* with companding, both as a function of the input level.

[3] J. F. Schouten, F. de Jager and J. A. Greefkes, Delta modulation, a new modulation system for telecommunication, Philips tech. Rev. **13**, 237-245, 1951/52.

[4] J. A. Greefkes and K. Riemens, Code modulation with digitally controlled companding for speech transmission, Philips tech. Rev. **31**, 335-353, 1970.

Voice switch

The interruption of the closed electro-acoustic circuit needed to prevent feedback on the loudspeaking telephone is controlled by a comparison of the strength of the outgoing and incoming signals in the two sets. The digitally controlled companding offers an excellent facility for this purpose, making it only necessary to compare the two control voltages v_c from the transmitter and receiver. This is done in a differential amplifier. The output voltage from the amplifier controls a switch that during each picture line passes either the eight sound bits or the standard series of alternating '0's and '1's (corresponding to a d.c. voltage of zero for the delta demodulator).

Measures have to be taken to prevent too frequent switching. The sound from the loudspeaker is always louder than that of a natural speaking voice. So much crosstalk can occur on the microphone of the subscriber who is not speaking as to cause his voice switch to close and consequently that at the other end to open. The echo in the room at the receiving end can also produce a microphone signal when a pause occurs in the received signal. To ensure that switchover does not occur too readily or too rapidly, the voice switch does not react until the microphone signal has been dominant for at least 40 milliseconds. In addition, the control voltage derived from the transmitter is attenuated by a factor of 2; see *fig. 8*. For a connection to be made, a sound must now be produced that is coded in steps twice as large as the received signal, which means that it must have almost twice the amplitude.

One disadvantage of this asymmetry in the control signals is that during breaks in the outgoing speech signal the control voltage of the receiver always predominates and the outgoing connection is easily interrupted. To prevent this a hold circuit is incorporated, whose function is to ensure that a d.c. voltage V_h is subtracted from the control signal from the receiver when the voice switch is closed (fig. 8).

If the ordinary handset is used, the voice switch is short-circuited; there is then an uninterrupted connection.

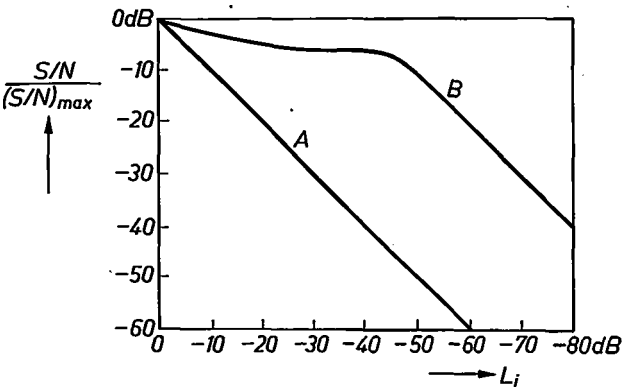


Fig. 7. Ratio of signal S to modulation noise N as a function of the input level L_i. Curve A: ordinary delta modulation. Curve B: delta modulation with digitally controlled companding.

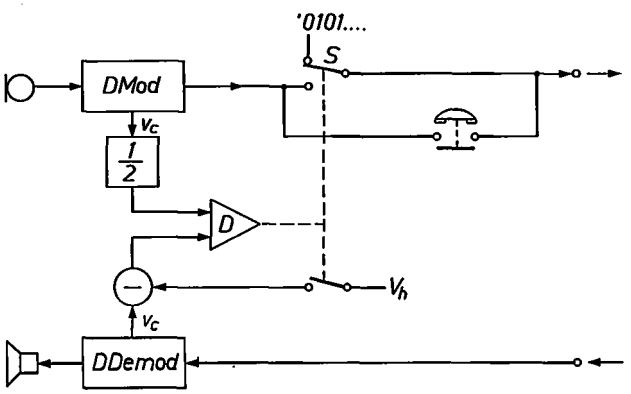


Fig. 8. Voice switch S in the outgoing sound channel is controlled by the differential amplifier D, which compares the control signals v_c from the delta modulator *DMod* and delta demodulator *DDemod*. To ensure more reliable operation the control voltage at the transmitting end is reduced by a half and a constant amount V_h is subtracted from that at the receiving end when the voice switch is closed.

Possibilities of other applications

A welcome additional facility is the conference circuit, i.e. a simultaneous interconnection between three or more video-telephone sets. In the conference mode the otherwise independent bit-clock oscillators in the transmitters of these sets are all synchronized with a single central oscillator. A voice-actuated switch ensures that the picture of the participant who is speaking appears on all the other screens. Use of the conference circuit necessitates a special conference unit which contains a number of auxiliary circuits and is best installed in the exchange.

The extra signalling capacity of the digital videotelephone system described above will prove of great value in this kind of application in the future. More than four line periods in the field-blanking interval are available for the transmission of additional information; see fig. 2. It is also conceivable that once the receivers have been synchronized the first five lines of the field-blanking interval could be used in whole or part for information other than the continuous '0101' series.

The extra capacity might be used to eliminate the objectionable phase jump that occurs on switching over to another participant. To achieve this result the phase of the picture to which the change-over has to be made would have to be registered just when the current picture phase was reaching zero, and passed on to the

participants before the change-over. The change-over would then take place at the next zero crossing and the counters in the receivers would be set not to zero but to the phase that had just been indicated. Special arrangements are generally necessary in the monitor to ensure that the field deflection can follow the discontinuity without any difficulty.

The extra signalling capacity might also be used to pass on messages from the participants to the conference unit. A participant might, for example, ask for a picture that was different from the one allocated by the voice-actuated switch.

Conference telephony is by no means the only extra facility that might be added to the video-telephone system. For example, a video-telephone set already contains many of the elements needed in a computer video terminal, such as the digital connection with bit-clock synchronization and the screen. The signals from an added keyboard could be accommodated in signalling time not as yet fully occupied. Another possibility is the use of the wideband video-telephone circuits for the transmission of data and facsimile.

Summary. There are various possible ways of arranging a video-telephone system, with or without separate transmission lines for sound and signalling. Separate lines are not required in the experimental system at Philips Research Laboratories. Sound and signalling are digitally coded and transmitted during the blanking intervals of the video signals. Each set has two bit-clock oscillators (1.0016 MHz); that of the receiver is synchronized by the other party's transmitter. Numerous control signals, including those for picture scanning in camera and monitor, are derived by frequency division from these oscillators. Series of '0101's together with a synchronization word in each field are used for synchronization; the word is detected by a matched filter. Delta modulation with digitally controlled companding is used for coding the speech signals. The control signal employed in this process also controls the voice switch; this prevents oscillation due to electro-acoustic feedback. Other potential applications include conferences taking in more than two sets and use as a computer video terminal.

Recent 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*
Mullard Research Laboratories, Redhill, Surrey, England	*M*
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*
MBLE Laboratoire de Recherches, 2 avenue Van Becelaere, 1170 Brussels (Boitsfort), Belgium	*B*
Philips Laboratories, 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A. (by contract with the North American Philips Corp.)	*N*

Reprints of most of these publications will be available in the near future. Requests for reprints should be addressed to the respective laboratories (see the code letter) or to Philips Research Laboratories, Eindhoven, The Netherlands.

H. Bex: Automatic measurement of crossmodulation.
Nachrichtentechn. Z. **28**, 327-329, 1975 (No. 9). *A*

K. Board, J. M. Shannon & A. Gill: C.C.F.E.T.: an active charge-coupled device.
Electronics Letters **11**, 452-453, 1975 (No. 19). *M*

G.-A. Boutry: Photoémission et vision nocturne.
Sciences et Techniques No. 18, 7-11, 1974. *L*

P. Branquart, J.-P. Cardinael, J. Lewi, J.-P. Delescaille & M. Vanbegin: Data structure handling in ALGOL 68 compilation.
Proc. Int. Conf. on ALGOL 68, Winnipeg 1974, pp. 59-80. *B*

J. W. Broer: Your interim report — did you time it rightly?
J. tech. Writ. Comm. **5**, 11-15, 1975 (No. 1). *E*

H. H. Brongersma: Structuuronderzoek van vastestof-oppervlakken.
Chem. Weekblad **71**, No. 12, 16-20, 1975. *E*

K. H. J. Buschow, J. F. Olijhoek & A. R. Miedema: Extremely large heat capacities between 4 and 10 K.
Cryogenics **15**, 261-264, 1975 (No. 5). *E*

K. L. Bye: Optische Beobachtung ferroelektrischer Domänen in elektrooptischer PLZT-Keramik.
Ber. Dtsch. Keram. Ges. **52**, 234-235, 1975 (No. 7). *M*

F. Caillaud, G. Durand, C. Le Can & G. Le Floch: Circuits intégrés pour télécommande en technologie C-MOS.
C.R. Coll. Int. sur les Circuits intégrés complexes, Paris 1974, pp. 295-303. *L*

F. M. A. Carpay, S. Mahajan (Bell Laboratories, Murray Hill, N.J.), **G. Y. Chin** (Bell Labs., M.H.) **& J. J. Rubin** (Bell Labs., M.H.): Slip-induced cracking in molybdenum single crystals.
Scripta metall. **9**, 451-457, 1975 (No. 5). *E*

T. A. C. M. Claasen, W. F. G. Mecklenbräuker & J. B. H. Peek: Remarks on the zero-input behavior of second-order digital filters designed with one magnitude-truncation quantizer.
IEEE Trans. **ASSP-23**, 240-242, 1975 (No. 2). *E*

G. Clément, C. Loty, J. P. Roux & C. Chancel (Commissariat à l'Energie Atomique, Limeil): The design of a new electron optics for a picosecond streak camera.
Proc. 11th Int. Congress on High speed photography, London 1974, pp. 130-135; 1975. *L*

J. Cornet & D. Rossier: Les verres semi-conducteurs: dispositifs nouveaux et technologie.
Coll. Int. sur les Matériaux pour les composants électroniques, Paris 1975, pp. 194-202. *L*

P. Devijver: Sur la signification de certains critères de l'analyse discriminante appliqués au problème de la reconnaissance des formes.
C.R. Séminaire IRIA Classification automatique et perception par ordinateur, 1974, pp. 285-304. *B*

H. Durand: Solar energy, from science fiction to reality? or: Acta Electronica, 15 years later.
Acta Electronica **18**, 261-266, 1975 (No. 4). *(Also in French.)* *L*

F. L. Engel (Institute for Perception Research, Eindhoven): Visibility, conspicuousness and attention.
Ophthalmologica **171**, 41-42, 1975 (No. 1).

D. den Engelsen & E. P. Honig: Bouwen met monomoleculaire lagen.
Chem. Weekblad **71**, No. 12, 28-32, 1975. *E*

G. Eschard: The microchannel plates: a survey of the possibilities of channel multipliers arrays in nano and picosecond detection and imaging.
Proc. 11th Int. Congress on High speed photography, London 1974, pp. 163-169; 1975. *L*

E. Fabre & M. Mautref: Caractérisation physique des cellules solaires au silicium par l'étude des réponses spectrales.
Acta Electronica **18**, 331-338, 1975 (No. 4). *L*

A. J. Fox & T. M. Bruton: Electro-optic effects in the optically active compounds $Bi_{12}TiO_{20}$ and $Bi_{40}Ga_2O_{63}$.
Appl. Phys. Letters **27**, 360-362, 1975 (No. 6). *M*

M. J. C. van Gemert & A. Suggett (Unilever Research Colworth/Welwyn, Sharnbrook, Beds., England): Multiple reflection time domain spectroscopy, II. A lumped element approach leading to an analytical solution for the complex permittivity.
J. chem. Phys. **62**, 2720-2726, 1975 (No. 7). *E*

J. J. Goedbloed: Investigation of parasitic oscillations in IMPATT-diode oscillators by a simple locus chart.
Electronics Letters **11**, 54-56, 1975 (No. 3). *E*

J. M. Goethals & J. J. Seidel (Eindhoven University of Technology): The regular two-graph on 276 vertices.
Discrete Math. **12**, 143-158, 1975 (No. 2). *B*

P. Hansen, J. Schuldt, B. Hoekstra & J. P. M. Damen: Anisotropy and magnetostriction of ruthenium-substituted lithium ferrite and nickel ferrite.
Phys. Stat. sol. (a) **30**, 289-298, 1975 (No. 1). *H, E*

D. Hennings: Mikrostruktur und mechanische Stabilität von dotierten Bleititanat-Keramiken.
Ber. Dtsch. Keram. Ges. **52**, 220-222, 1975 (No. 7). *A*

W. J. van den Hoek & G. Rouweler (Philips Lighting Division, Eindhoven): On thermodynamic calculations of chemical transport in halogen incandescent lamps.
Philips Res. Repts. **31**, 23-34, 1976 (No. 1).

W. K. Hofker, D. P. Oosthoek, N. J. Koeman (all with Philips Research Labs., Amsterdam Division) **& H. A. M. de Grefte:** Concentration profiles of boron implantations in amorphous and polycrystalline silicon.
Radiation Effects **24**, 223-231, 1975 (No. 4). *E*

E. Kauer, R. Kersten & F. Mahdjuri: Photothermal conversion.
Acta Electronica **18**, 295-304, 1975 (No. 4). *A*

E. T. Keve: Röntgenuntersuchungen zu Phasenumwandlungen in PLZT-Keramik.
Ber. Dtsch. Keram. Ges. **52**, 232-234, 1975 (No. 7). *M*

F. M. Klaassen: Device physics of integrated injection logic.
IEEE Trans. **ED-22**, 145-152, 1975 (No. 3). *E*

F. M. Klaassen: A MOS model for computer-aided design.
Philips Res. Repts. **31**, 71-83, 1976 (No. 1). *E*

F. M. Klaassen, W. de Groot & F. L. van de Markt: Computer algorithm to determine MOS process parameters.
Philips Res. Repts. **31**, 84-92, 1976 (No. 1). *E*

W. L. Konijnendijk & J. H. J. M. Buster: Raman-scattering measurements of arsenic-containing oxide glasses.
J. non-cryst. Solids **17**, 293-297, 1975 (No. 2). *E*

J.-P. Krumme, P. Hansen, G. Bartels & D. Mateika: Control of the uniaxial magnetic anisotropy in LPE-grown iron garnet films by small ruthenium substitutions.
J. appl. Phys. **46**, 2801-2803, 1975 (No. 6). *H*

J.-P. Krumme, B. Hill, J. Krüger & K. Witter: A highly sensitive reversible and nonvolatile hybrid photoconductive/magneto-optic storage material.
J. appl. Phys. **46**, 2733-2736, 1975 (No. 6). *H*

J. Lemmrich: Winkelregelungen von Gleichstrommaschinen mit frequenzanaloger Signalverarbeitung.
Elektrotechn. Z. B **27**, 401-404, 1975 (No. 15). *H*

J. Michel & A. Mircea: Simulation de cellules solaires au silicium et comparaison avec des résultats expérimentaux.
Acta Electronica **18**, 311-330, 1975 (No. 4). *L*

M. Monneraye: Conducteurs au cuivre sérigraphiés en couches épaisses.
Coll. Int. sur les Matériaux pour les composants électroniques, Paris 1975, p. 132. *L*

B. H. Newton: M.I.C. TRAPATT oscillator for efficient S band operation.
Electronics Letters **11**, 299-300, 1975 (No. 14). *M*

A. E. Pannenborg (Philips Board of Management, Eindhoven): Technology push versus market pull — the designer's dilemma.
Electronics & Power **21**, 563-566, 1975 (15 May).

L. J. van der Pauw: The radiation and propagation of electromagnetic power by a microstrip transmission line.
Philips Res. Repts. **31**, 35-70, 1976 (No. 1). *E*

G. Piétri: Design and performances of fast phototubes for photon counting or for high speed imaging.
Proc. 6th Int. Symp. IMEKO TC on Photon detectors, Siofok (Hungary) 1974, pp. 142-161. *L*

R. Polaert & J. Rodière: Improvement of the performances of high speed cinematography through the use of a proximity focusing microchannel image intensifier.
Proc. 11th Int. Congress on High speed photography, London 1974, pp. 170-177; 1975. *L*

L. J. Poldervaart (Eindhoven University of Technology) **& C. G. Sluijter:** Photonics: the profile of a new discipline.
Proc. 11th Int. Congress on High speed photography, London 1974, pp. 585-589; 1975. *E*

W. Rey: On least p-th power methods in multiple regressions and location estimations.
BIT **15**, 174-184, 1975 (No. 2). *B*

C. J. M. Rooymans: Recent trends in magnetic bubble devices.
Rev. Phys. appl. **10**, 179-181, 1975 (No. 3). *E*

J. A. J. Roufs (Institute for Perception Research, Eindhoven): The Standard Observer: a controversial subject.
Ophthalmologica **171**, 43-44, 1975 (No. 1). *E*

J. M. S. Schofield: The physics of gas discharge display cells.
Proc. 12th Int. Conf. on Phenomena in ionized gases, Eindhoven 1975, Part 1, p. 73. *M*

P. J. Severin, H. Bulle, G. Poodt & J. D. Wasscher: On the technique and evaluation of angle-beveling silicon epitaxial layers.
J. Electrochem. Soc. **122**, 440-443, 1975 (No. 3). *E*

J. Smith: Experimental storage display panels using dc gas discharges without resistors.
IEEE Trans. **ED-22**, 642-649, 1975 (No. 9). *M*

M. J. Sparnaay: Oppervlakteonderzoek bij Philips.
Chem. Weekblad **71**, No. 12, 13-15, 1975. *E*

J. L. Spoormaker: Design of reliable plastic assemblies.
Proc. 1975 Ann. Reliability and Maintainability Symp., Washington D.C., pp. 498-503. *E*

A. L. N. Stevels & A. D. M. Schrama-de Pauw: On the luminescence, excitation and UV-absorption spectra of CsI activated with Na.
Philips Res. Repts. **31**, 1-22, 1976 (No. 1). *E*

T. Thalhammer: Future energy demand and the role of solar energy.
Acta Electronica **18**, 267-273, 1975 (No. 4). *E*

J. A. T. Verhoeven & J. J. Vrakking: Analyse van oppervlakken met Auger-elektronenspectroscopie.
Chem. Weekblad **71**, No. 12, 20-24, 1975. *E*

J. M. P. J. Verstegen, D. Radielović & L. E. Vrenken (Philips Lighting Division, Eindhoven): A new generation deluxe fluorescent lamp — combining an efficacy greater than 80 lm/W with a color rendering index of about 85.
J. Illum. Engng. Soc. **4**, 90-98, 1975 (No. 2).

H. J. Vink: Food-production and bioenergy.
Acta Electronica **18**, 305-310, 1975 (No. 4). *E*

H. W. Werner & G. E. Thomas: Emissie van ionen en van licht door beschieting van oppervlakken met snelle ionen: SIMS en BLE.
Chem. Weekblad **71**, No. 12, 24-28, 1975. *E*

W. Yu, R. R. Alfano (both with City College of New York), **C. L. Sam & R. J. Seymour:** Spectral broadening of picosecond 1.06 μ pulse in KBr.
Optics Comm. **14**, 344-347, 1975 (No. 3). *N*

Contents of Philips Telecommunication Review **34**, No. 1, 1976:

F. L. van den Berg: The Philips Doppler VOR beacon RN 200 (pp. 1-10).
F. L. Jansen & G. W. Versteeg: The quality of reed contact units (pp. 11-21).
W. F. Njio & C. J. Verkooijen: Frequency generating equipment for FDM transmission systems (pp. 22-37).
C. A. Schipper: Thin film circuits in transmission equipment (pp. 38-41).
A. Potuit: Belgium chooses Philips coaxial line equipment for 60 MHz network (p. 42).

Contents of Electronic Applications Bulletin **33**, No. 3, 1976:

Switched-mode power supply transformer design nomograms (pp. 95-121).
Transformer and choke design for forward-converter switched-mode power supplies (pp. 125-143).
Components for switched-mode power supplies (pp. 145-147).

Contents of Mullard Technical Communications **13**, No. 129, 1976:

F. J. Burgum: Switched-mode power supply transformer design nomograms (pp. 354-378).
M. C. Basell: 50W multiple-output switched-mode power supply (pp. 379-387).
CL8880 series radar traffic sensors (pp. 388-392).

Contents of Valvo Berichte **20**, No. 1, 1976:

G. Tauchen: Bildröhrenheizung aus der Horizontalendstufe (pp. 1-12).
J. Koch: Permanentmagnetische Polfühligkeit in Gleichstrommotoren mit Ferroxdure-Segmenten (pp. 13-22).
J. Koch: Entmagnetisierungsenergie und Anziehungskraft von Permanentmagneten (pp. 23-32).
W. Hetterscheid: Schaltverhalten bei Transistoren für hohe Spannungen (pp. 33-44).

PHILIPS TECHNICAL REVIEW

VOLUME 36, 1976, No. 9

Apparatus for solid-state research at very high pressures

M. Brouha and A. G. Rijnbeek

A rapidly growing interest in the use of very high pressures developed after 1955, when success was first obtained in the United States and in Sweden in producing synthetic diamond. The stimulus that this provided for the use of very high pressures resulted in the development of new high-pressure apparatus, both for synthesis and for research. The article below describes apparatus for solid-state research, with emphasis on recent developments at Philips Research Laboratories.

Introduction

There are many areas of materials research in which the use of high pressures helps to provide information that can often not be obtained by any other means. Yet high-pressure techniques are not in fact very widely used, probably because of a lack of familiarity with them. It could also seem that work with high pressures was very difficult, although in reality it is no more complicated than working in cryogenics, to quote one example.

The applications of high pressures are of two kinds: in the preparation [1] of materials, and in their investigation [2].

Examples of the first type are the high-pressure synthesis of polyethylene, the conversion of graphite into diamond and the conversion of hexagonal into cubic boron nitride. Almost all solid substances under pressure go through reversible changes of state, usually to a denser phase. In some cases, e.g. diamond, these phases are metastable or even stable, so that they persist after the temperature and pressure treatment. Another application is in the high-pressure sintering and compaction of tungsten, tungsten carbide and ferrites.

In materials research it is of interest to find out how a particular property of the material varies with changes in volume, obtained by the application of high pressures.

As we shall see, it is relatively simple to perform effective measurements in the pressure range up to 100 kbar (1 bar = 0.9869 atm), since purely hydrostatic and accurately measurable pressures can be used in that range.

In solids a pressure of 100 kbar causes a change in volume ranging from a few per cent (for diamond, which has the lowest compressibility) to some 30 or 40% (for the alkali metals) [3]. For most materials the change in volume is about 10%, which means that the interatomic distances can be reduced by several per cent. Many properties of solids depend closely on the interatomic distance. A pressure of 100 kbar can therefore bring about considerable changes in those properties.

The requirements to be met are not the same for preparative apparatus as for apparatus for investigatory purposes. In synthesis the most important feature

Ir M. Brouha is with Philips Research Laboratories, Eindhoven; Ir A. G. Rijnbeek, formerly with Philips Research Laboratories, is now with the Philips Audio Division, Eindhoven.

[1] See for example:
R. S. Bradley and D. C. Munro, High pressure chemistry, Pergamon Press, Oxford 1965;
K. E. Weale, Chemical reactions at high pressures, Spon, London 1967;
J. B. Goodenough, J. A. Kafalas and J. M. Longo, High-pressure synthesis, and C. J. M. Rooymans, High-pressure techniques in preparative chemistry, in: P. Hagenmuller (ed.), Preparative methods in solid state chemistry, Academic Press, New York 1972, pp. 1-69 and 71-131 respectively.

[2] See for example C. C. Bradley, High pressure methods in solid state research, Butterworth, London 1969, and G.C. Ulmer (ed.), Research techniques for high pressure and high temperature, Springer, Berlin 1971.

[3] K. A. Gschneidner, Jr., Physical properties and interrelationships of metallic and semimetallic elements, Solid State Physics 16, 275-426, 1964.

is that the desired temperature and pressure range should be reached in the largest possible volume, while any inhomogeneities are of secondary importance. In materials research, however, temperature and pressure have to be well defined and accurately measurable. Almost all the apparatus described in this article was designed with our own materials research in mind. Three types of apparatus may be distinguished: the simple pressure vessel, suitable for pressures up to 20 kbar; the compound pressure vessel, which can be used up to 45 kbar; and finally the Bridgman anvil, with which pressures of several hundred kbar can be attained. In all cases the specimen is surrounded by a medium that transmits the pressure in the vessel or anvil to the specimen.

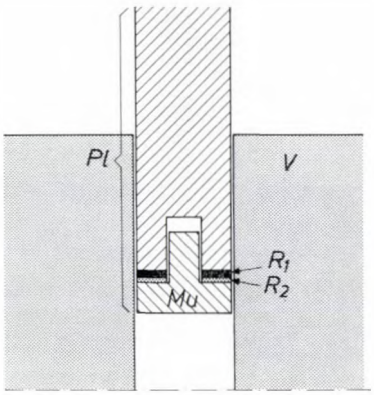


Fig. 1. Bridgman seal in a simple pressure vessel. The seal in the gap between the wall *V* and the plunger *Pl* is formed by a rubber ring R_1 and a mild-steel ring R_2, which are compressed between the 'mushroom' plug *Mu* and the end of the plunger.

Pressure media

In materials research it is desirable that the pressure should be distributed isotropically over the specimen. This can be assured by using a gas or a liquid as the medium for transmitting the pressure to the specimen. A practical difficulty here is that most liquids solidify fairly rapidly with increasing pressure; for example high-pressure oil solidifies at 10 kbar and pentane or petroleum ether solidify at 40 kbar. For pressures greater than 40 kbar a 4 : 1 mixture of methanol and ethanol is used; this mixture remains liquid up to 100 kbar. The most suitable gas for a pressure medium is helium, which does not solidify at room temperature until the pressure reaches 120 kbar. If the temperature is decreasing, this is the substance that remains liquid longest at ordinary pressures; at 4.2 K helium changes to the solid state at a pressure of 140 bars.

Beyond the limits stated above experiments under purely hydrostatic conditions are in fact impossible. When it is necessary to work at pressures greater than 100 kbar or at low temperatures, the materials used for pressure media are solids that readily undergo plastic deformation when subjected to a pressure gradient, thus behaving 'quasi-hydrostatically'. Examples of such substances are hexagonal boron nitride, silver chloride and pyrophyllite (a natural hydrous aluminium silicate found as a mineral). At extremely low temperatures solid helium is also a very good pressure medium because of the weak bond between its atoms.

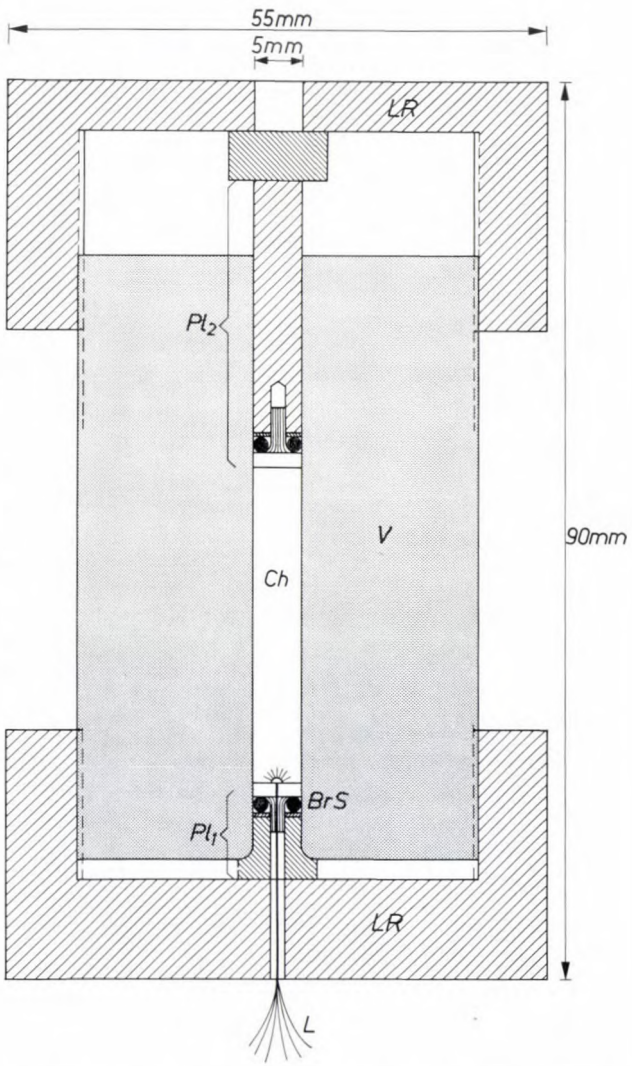


Fig. 2. Simple pressure vessel of maraging steel or beryllium copper for measurements up to 20 kbar. After the force has been applied the upper locking ring *LR* is tightened and the vessel can be removed from the press without affecting the pressure inside. The electrical leads *L* pass through the fixed plunger Pl_1 with its Bridgman seal *BrS*. The specimen is contained in the chamber *Ch*. *V* is again the wall of the vessel.

Simple pressure vessels

In the simplest method of applying pressure the specimen is placed in a cylinder filled with liquid. The pressure is applied by exerting a force on a plunger that seals off one end of the cylinder. This kind of pressure vessel is used in many versions and in differing dimensions. The vessels are made of a good steel (often of

'maraging steel', because of its high tensile strength, the ease with which it can be hardened, and its high dimensional stability after hardening), with an external diameter five to seven times the internal diameter. This kind of vessel can be used at pressures up to 20 kbar. In the design stage it is necessary to bear in mind that a liquid medium such as pentane or petroleum ether will be compressed at 20 kbar to only two-thirds of the original volume. The device used to seal the plunger, a 'Bridgman seal' [4], is shown schematically in *fig. 1*. The device is self-sealing; the 'mushroom' plug *Mu* of the plunger *Pl* makes the pressure in the sealing rings R_1 and R_2 higher than that in the liquid. As the pressure is raised the sealing is at first provided by the rubber ring R_1. At pressures of several kbar the seal is obtained by flow in the mild-steel ring R_2. An O-ring has the advantage of low friction but can only be used up to 10 kbar. Fixed Bridgman seals are also used. The vessel is then brought up to pressure via a capillary tube. This method is mainly used with gases on account of their high compressibility, to avoid having to use very long vessels with very long plungers. The pressure can then be applied in two or more stages. The final stage in such a pump is however again a cylinder and plunger, fitted with a Bridgman seal.

To enable measurements to be made on the sample, various provisions have to be made. In many cases it is desirable to be able to mount the pressure vessel in the measurement arrangement away from the press. Most of our vessels are therefore provided with locking rings *LR*, which can be used to hold the plungers in position after setting to the desired pressure in the press; see *fig. 2*. If the pressure has to be altered, the vessel is placed in the press again. The force required to produce a pressure of 20 kbar is about 40 kN (4 tonnes) for the vessel illustrated in fig. 2. This force can be obtained with quite simple arrangements.

Pressure vessels that can be removed from the press for measurements can be very small (external diameter about 5 mm, length about 2 cm). They can be placed in an instrument such as a magnetometer, or placed under a microscope, although in that case windows must be provided in the vessel. The vessels are suitable for measurements over a wide temperature range (4-500 K).

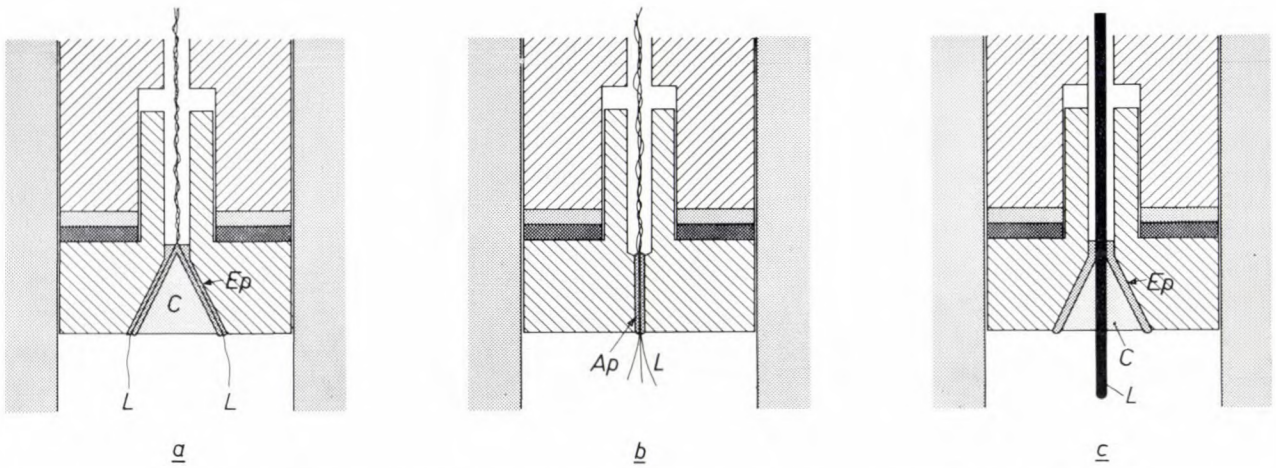


Fig. 3. *a*) A frequently used arrangement for taking electrical leads through a plunger with a Bridgman seal. The leads *L* are encapsulated in epoxy resin *Ep* around the cone *C*. *b*) Simplified and improved arrangement for the electrical leads. A central aperture *Ap* is produced by micro-spark machining. The diameter of the aperture is so small that the leads, which are encapsulated in epoxy resin, are not pressed outwards. *c*) Lead-through for a very thick wire. The wire is soldered into a central hole in a cone *C*, which is insulated from the plunger with epoxy resin.

Electrical connections

In high-pressure vessels the electrical connections required for measurements are a problem in themselves. The lead-throughs have to be designed in such a way that the vessel cannot leak and the wires are not damaged by pressure gradients. One widely used solution is illustrated in *fig. 3a*. The leads are encapsulated in epoxy resin between cone *C* and the conical aperture in the plunger [2]. The reliability of these leads depends on factors such as the wire diameter, the apex angle and the height of the cone.

We use a simpler and, in our experience, a more reliable method of taking thin wires through the wall (fig. 3*b*). A central aperture *Ap* is produced in the plunger by micro-spark machining; the lower part has

[4] P. W. Bridgman, The physics of high pressure, Bell, London, and Dover Publ., New York, 1970.

a diameter of 0.3 mm and a length of 5 mm. This diameter is large enough to take four 0.1-mm wires or ten 0.05-mm wires, embedded in epoxy resin. The number of wires can be increased by simply providing more holes in the plunger. Thick wires, which might be necessary for a magnet coil or a heating element, are separately soldered in a central aperture in a cone; see fig. 3*c*. A 'Thermocoax' thermocouple is used for temperature measurement. Its leads are insulated from the stainless-steel casing, which is soldered into a hole in the 'mushroom'.

Measurements on specimens in simple pressure vessels

Electrical measurements

The pressure vessels with the connections described above allow electrical measurements to be made on specimens at various temperatures and pressures.

The effects of pressure on the Hall effect and the magnetoresistance effect are usually measured by means of a magnetic field excited outside the vessel. This must then of course be made of non-magnetic material. It is also possible to produce a magnetic field by means of a magnet coil inside the vessel.

The susceptibility of a material is found by measuring the inductance of a coil or the mutual inductance of a transformer whose core incorporates the specimen.

The compressibility of a material can be studied with the aid of strain gauges. These are attached to the specimen to be measured and to a reference specimen. The resistances of the gauges are compared in a Wheatstone bridge.

Magnetization measurements

Magnetization measurements can be made on the complete vessel containing the specimen, provided that the magnetic contribution from the vessel is sufficiently small. This condition can be met by making all parts of the vessel of a non-magnetic material such as beryllium copper (a copper alloy containing about 2% of beryllium). Beryllium copper can be used up to a pressure of about 15 kbar; beyond that value plastic deformation sets in.

The vessel used for magnetization measurements has to be small because the distance between the magnet poles in the arrangement is limited, and it must also be possible to construct a furnace or cryostat around it. A diagram of a vessel that we have designed for such measurements is shown in *fig. 4*. It is difficult to use a Bridgman seal in a vessel of such a small internal diameter (1.2 mm), but in fact it is not necessary. We have been able to produce an adequate seal simply with a rubber plug *Plug* and an unhardened beryllium-copper disc *D*.

Optical measurements

The windows of pressure vessels for optical measurements are generally made of sapphire and set in Bridgman seals (*fig. 5*). Good adhesion between the sapphire and the metal is obtained by making use of Van der Waals forces between two smoothly polished surfaces. The vessels are sufficiently small to allow observations to be made under a microscope.

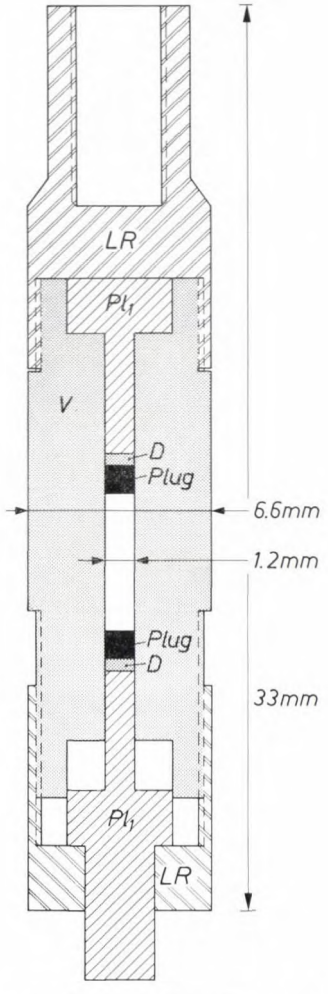


Fig. 4. A non-magnetic vessel of beryllium copper for measuring magnetization under pressure. The vessel is small enough to enable it to be enclosed inside a furnace or cryostat and placed between the poles of a magnet. The seal consists of a rubber plug *Plug* and a disc *D* of annealed BeCu. Other symbols have the same significance as in the previous figures.

Compound vessels

The vessels used for pressures greater than 20 kbar are *compound* vessels. These consist of an inner cylinder clamped inside one or more outer cylinders. The inner cylinder is prestressed by means of a shrink fit or a tapered structure. *Fig. 6* shows a compound steel vessel with a taper fit. As in simple vessels, the pressure me-

dium is a liquid and Bridgman seals are used. Steel vessels can be employed up to 30 kbar.

The pressure range can be extended to 45 kbar by making the inner cylinders and the plungers of tungsten carbide (also known as 'hard metal'), an alloy that has a very high compressive strength (55 kbar). Above 45 kbar the plungers are susceptible to plastic deformation. Tungsten-carbide cylinders are also prestressed axially with the aid of an extra hydraulic pressure cylinder. A special press is therefore necessary for working with these vessels.

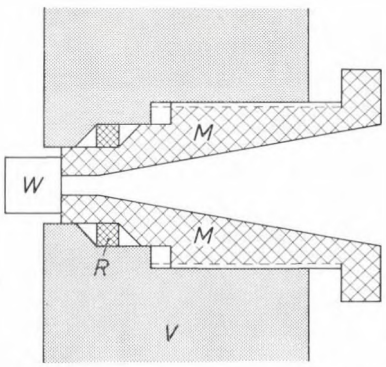


Fig. 5. Part of a wall of a pressure vessel for optical measurements; the sapphire window *W* is polished smooth and is fitted flat against the front end, also highly polished, of the mount *M*, which is sealed with a brass ring *R*.

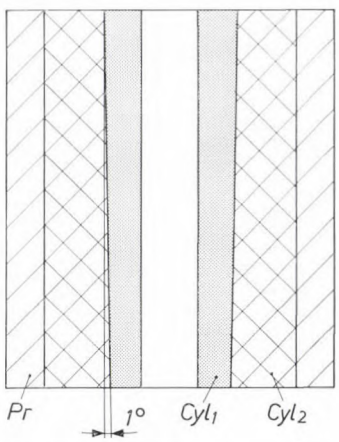


Fig. 6. Section through a compound vessel with an inner cylinder Cyl_1, which is prestressed by a tapered fit with the outer cylinder Cyl_2. It is surrounded by a mild-steel protecting ring *Pr*.

To avoid sealing problems with the plunger and the electrical leads solid substances are used as pressure media in hard-metal vessels. If it is however necessary to work under hydrostatic conditions, a PTFE ('Teflon') capsule filled with a liquid and completely enclosed in a solid pressure medium such as talc may be employed. A general view of such a vessel, as used by ourselves [5], is shown in *fig. 7a*. Fig. 7*b* gives a detailed cross-section of the pressure chamber with the PTFE capsule (Enc_1) and the other components. The capsule is filled with pentane or another fluid that remains liquid at very high pressures at room temperature. The seal around the lid of the capsule — an O-ring — and the electrical leads now only have to withstand the residual differential pressure in the solid medium over the length of the capsule (about 1 kbar).

The solution we have found for taking the electrical leads through the end plate (fig. 7*a*, *EP*) is one that takes up little room. The 'Thermocoax' thermocouple *Th* is taken centrally through the mild-steel cone, and the wires (up to eight) between two pyrophyllite cones. With this configuration (apex angle 36° and wall thickness of the pyrophyllite cones 0.4 mm) the wires are kept clear of each other and of the steel parts over the whole of the pressure range. The pressure change along the steel cone is so uniform that the wires are not affected by the very large difference in pressure.

The space available inside the capsule for the specimen *S* has a diameter of 12 mm and a height of 30 mm. At a pressure of 45 kbar the volume of the capsule is approximately halved; this is seen mainly as a reduction in height.

The graphite cylinder *F*, which is fitted around the capsule, is used as an electric furnace. The end plate is therefore electrically insulated from the vessel. By circulating liquid nitrogen through a cooling jacket around the vessel and using asbestos insulating plates between the vessel and the press the temperature can be varied between 100 and 600 K. Above 600 K the organic liquids used as pressure media start to decompose, as does also the PTFE. For experiments above 600 K boron nitride is used; the temperature can then be varied from 100 K to 1500 K. Because of the low thermal conductivity of pyrophyllite the temperature of the vessel can be kept sufficiently low by cooling with water or liquid nitrogen.

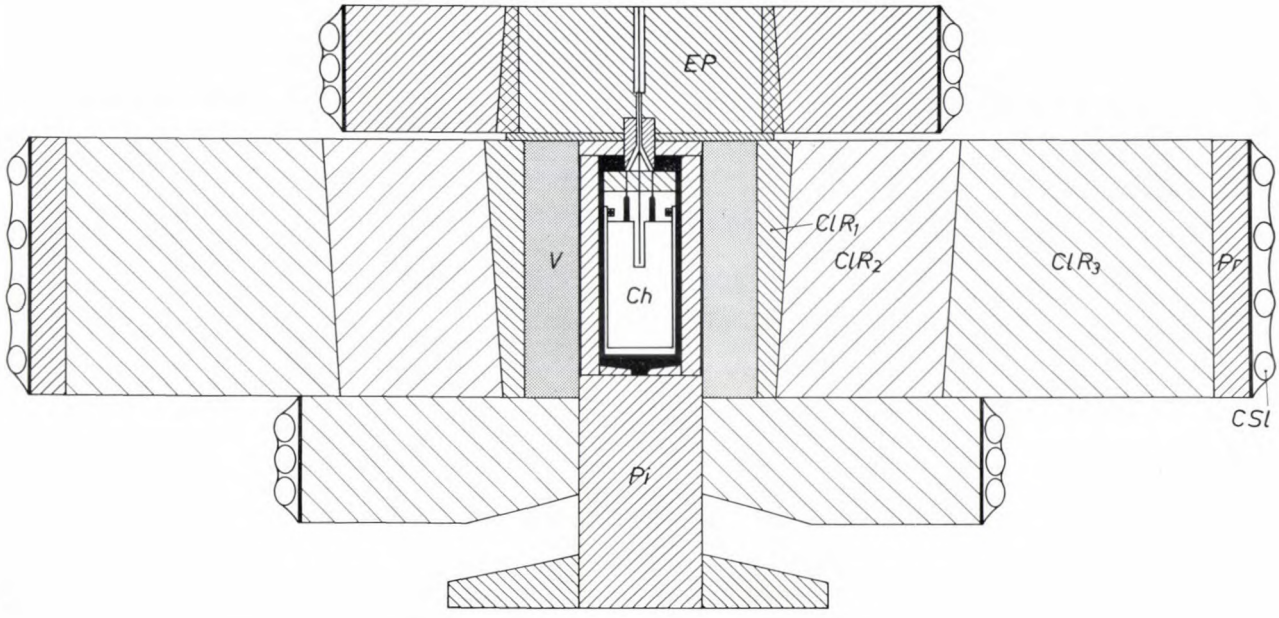
In this equipment we have determined quantities such as the pressure dependence of the electrical resistance of materials up to very high temperatures and the pressure dependence of the Curie temperature of ferromagnetic materials [6]. The Curie-temperature determinations were effected by measuring the inductance of a coil in which the specimen formed the core at a number of pressures and temperatures. The

[5] M. Brouha and A. G. Rijnbeek, Rev. sci. Instr. **44**, 852, 1973.

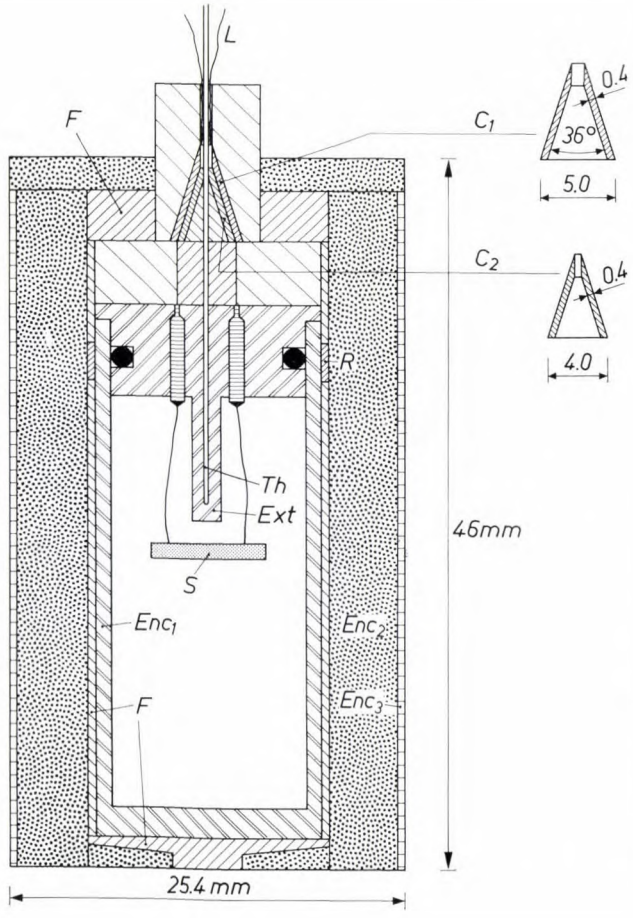
[6] M. Brouha and K. H. J. Buschow: J. appl. Phys. **44**, 1813, 1973; J. Physics F **3**, 2218, 1973; J. Physics F **5**, 543, 1975; J. appl. Phys. **46**, 1355, 1975; EPS Conf. Abstr. **1A**, 26, 1975. M. Brouha, K. H. J. Buschow and A. R. Miedema, IEEE Trans. **MAG-10**, 182, 1974.
J. W. M. Biesterbos, M. Brouha and A. G. Dirks, AIP Conf. Proc. **29**, 184, 1976.

susceptibility measured in this way for a small external field falls sharply at the Curie temperature.

Special modifications to the piston, such as making it conical, or using a number of pistons (four, five or eight) supporting each other and shaped to enclose a space allow the tungsten carbide to be exposed to a pressure of more than 55 kbar and the range to be extended to 80 kbar in relatively large volumes. Electrical connections, however, are extremely difficult to arrange here, and we do not therefore use this technique for electrical measurements, but change over to Bridgman anvils instead.



a



b

Fig. 7. *a*) Section through a clamped vessel *V* with piston *Pi*, enclosed at the top by a clamped end plate *EP*. These three parts are made of tungsten carbide. Around the hardened-steel clamping rings *ClR* there is a mild-steel protecting ring *Pr* with a cooling sleeve *CSl*. The space between *V*, *Pi* and *EP* — the pressure chamber — is shown in detail in (*b*). Part *Ch* of this space is available for the specimen. The vessel can be used for pressures up to 45 kbar and up to very high temperatures (1500 K).

b) A detailed diagram of the pressure chamber. The specimen *S* is contained in a liquid-filled 'Teflon' capsule Enc_1. The electrical leads *L* are taken out between two cones *C* of pyrophyllite. Capsule Enc_1 has a 'Teflon' cover sealed by an O-ring, which is strengthened by a steel ring *R*. In the centre of the part *Ext* protruding inwards there is a thermocouple *Th*, whose lead is taken out centrally. Around the capsule there is a graphite furnace *F*, a talc-filled cavity Enc_2 and a lead jacket Enc_3.

Bridgman anvils

Extremely high pressures can be reached in a straightforward manner with Bridgman anvils [4]. They are consequently widely used in various forms.

A Bridgman anvil has the shape of a truncated cone. It is made of tungsten carbide or other material with high compressive strength. A thin disc of pyrophyllite or a material with similar properties is placed between two such anvils. *Fig. 8* shows the specimen *S* in a space machined from the centre of the pyrophyllite disc *D*. The conical shape of the anvils *An* ensures that pressure is exerted only on the centre; the rest of the material is only intended as a kind of massive support. For an apex angle of the cone of between 140° and 160° the central part is found to be able to withstand a load three times as great as that of a simple cylindrical anvil. If the tungsten-carbide anvils are also clamped, a load four times as great, i.e. up to 200 kbar, can be sustained.

The pyrophyllite is at first pressed out between the anvils as the force is increased until a state of equilibrium is reached. By that time a certain pressure profile has become established over the surface of the anvil, with the highest pressure at the centre, gradually decreasing to 1 bar at the edge. This pressure profile depends on the mechanical properties of the pyrophyllite, a material admirably suited to this application.

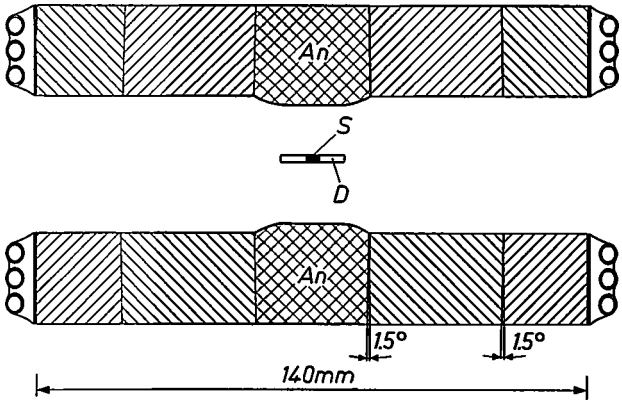


Fig. 8. Bridgman anvils. Between the clamped tungsten-carbide anvils *An* is a small pyrophyllite disc *D* containing the specimen *S*. A pressure of 200 kbar can be obtained with these anvils.

Because of its layered structure shearing depends very closely on the force applied perpendicularly to the direction of shear. In the state of equilibrium the pressure therefore increases extremely sharply from the edge towards the centre. If the disc is thin, the gradient is almost entirely confined to the edge and the pressure is constant over a large area. It is for that reason that the pyrophyllite discs used with Bridgman anvils generally have a diameter of about 5 mm and a thickness of 0.2 to 0.3 mm.

The discs that we use are of a thickness such that the pyrophyllite is almost but not quite pressed out from between the anvils. For relatively large anvils (with a surface diameter of 15 mm) this means that the disc is 1.5 mm high. Only in the immediate vicinity of the centre is the pressure constant. The area suitable for measurements is found to be 3 mm across. With these relatively large anvils the force necessary to obtain a particular pressure is little greater than with anvils of 5 mm diameter combined with a thin pyrophyllite disc. This is because the pressure in the area around the edge remains low and the pressure at the centre is three or four times the mean value.

In this way we have created sufficient room between the anvils to accommodate wires, a thermocouple and a furnace for measurements at various temperatures and up to very high pressures. The maximum attainable pressure is 200 kbar; when a number of wires and a furnace are used this value falls to 150 kbar, since the limit is determined by the properties of the pyrophyllite rather than those of the anvils. On increasing the force further there is very little increase in pressure at the centre; only the pressure profile changes.

At liquid-helium temperatures the same arrangement is used, but with slightly smaller anvils (with a surface of 10 mm diameter). A force of 500 kN (50 tonnes) is then sufficient to obtain maximum pressure. This force is provided by a double-acting toggle press in the lower part of the cryostat (*fig. 9*). With this arrangement the force is greatly amplified, so that the force on the toggle press remains below 10 kN. The walls of tubes T_1 and T_2 can therefore be made of thin stainless steel, which will give a low thermal conductivity.

An example of a configuration that we have used for the determination of Curie temperatures up to 150 kbar in a temperature range from 100 to 1250 K is shown in *fig. 10* [7]; the configuration includes pyrophyllite discs, furnace, wiring and thermocouple. The measurement leads $Wi_{1,2}$ are at different heights and form a transformer, with the specimen *S* as its core. The primary winding Wi_1 consists of a single turn around the specimen, plus a turn in the opposite sense to compensate for the magnetic contribution from the press. The secondary loop covers the complete surface of the two primary turns and also acts as a thermocouple (the chromel-alumel junction is located between the two halves of the specimen). The graphite discs $D_{4,5}$ above and below the specimen act as a furnace; the current also flows through the specimen itself. As for the tungsten-carbide vessels, refrigeration is provided by blowing liquid nitrogen through a cooling jacket surrounding the clamping rings. The temperature can be

[7] M. Brouha and A. G. Rijnbeek, High Temp. - high Press. **6**, 519, 1974.

reduced to 100 K. To determine the Curie temperature in this arrangement the secondary signal from the transformer is measured at various pressures and varying temperatures. *Fig. 11* shows an X-ray photograph of the centre of the apparatus after an experiment has been completed. The effect of the pressure can be seen clearly from the cracks in the pyrophyllite and the darker shading at the centre, where the pressure was highest.

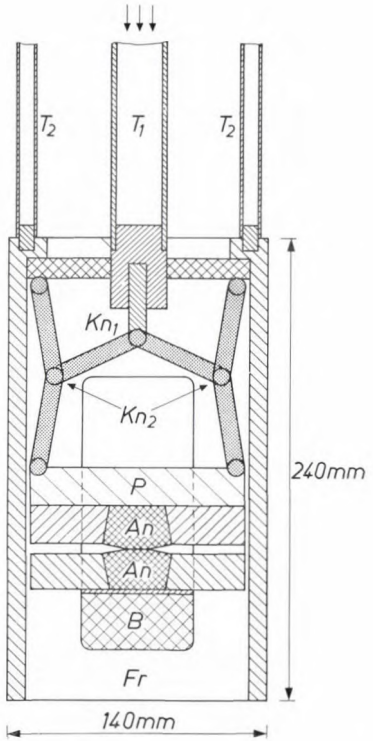


Fig. 9. A double-acting toggle press, used to apply force to a Bridgman anvil in a cryostat at liquid-helium temperature. The force is applied to the stainless-steel tube T_1 and very considerably amplified by being applied via the toggles Kn_1 and Kn_2 to the plate *P*. With this arrangement the actual force applied to the toggle press need only be small. This means that the tubes T_1 and T_2 can be thin-walled stainless-steel tubes, with low thermal conductivity. The Bridgman anvil rests on the beam *B*, which is inserted through the holes in the frame *Fr*. By placing one or more thin plates on top of *B* the anvil can be raised to the best position for operating the press.

A magnetic field of 6×10^5 A/m (about 8 kOe) can be obtained with a coil around the clamping rings at the same height as the specimen. Electrical resistance is generally measured with a similar arrangement, but with the wires in a single plane. They are simply laid over the specimen and the first axial pressure ensures a good contact.

The anvil material that has the highest compressive strength is diamond. The dimensions of available diamonds are rather small, so that in practice it is impossible to make anvils with an external diameter greater than 5 mm. Formerly the specimen was simply placed between the diamonds. This gave poor definition of the pressure, however. G. J. Piermarini and S. Block [8] have developed a better technique, whose principles we have adopted. An 'Inconel' ring with a hole in it of 0.2 mm diameter is placed between the diamonds. The specimen to be investigated, a liquid as the pressure medium and a second specimen for calibration of the pressure are located in the hole. The part of the ring between the diamonds becomes thinner with increasing pressure, because of plastic deformation. This produces an increase in pressure in the liquid. Extremely high pressures, up to 500 kbar, can be obtained with this

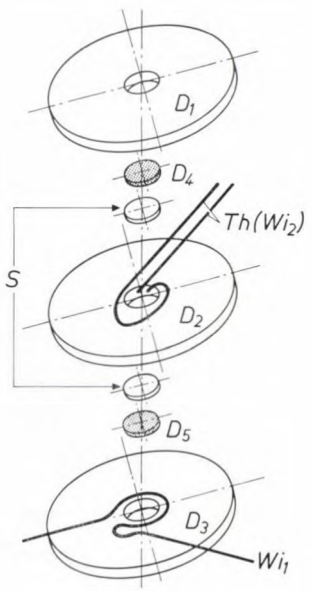


Fig. 10. An exploded view of the assembly used between the Bridgman anvils, which are used for determining Curie temperatures. Between the two halves of the specimen *S* there is a thermocouple *Th*, which also acts as the secondary winding Wi_2 of the transformer used for measuring the permeability of the specimen. The primary Wi_1 consists of one turn around the specimen and a turn in the opposite sense. This turn in the opposite sense provides compensation for the magnetic contribution from the press. The wiring is located in grooves machined in the pyrophyllite discs $D_{1,2,3}$. The graphite discs $D_{4,5}$ function as a furnace.

arrangement. *Fig. 12* shows our experimental arrangement. The specimen remains visible through the conical holes and the diamonds *An*, so that the apparatus can be used for optical experiments and X-ray diffraction.

Pressure measurement

The only method of measuring the pressure directly is to determine the force per unit surface area of the plunger. This is difficult, however, because of friction from the seal and elastic deformation in the plunger

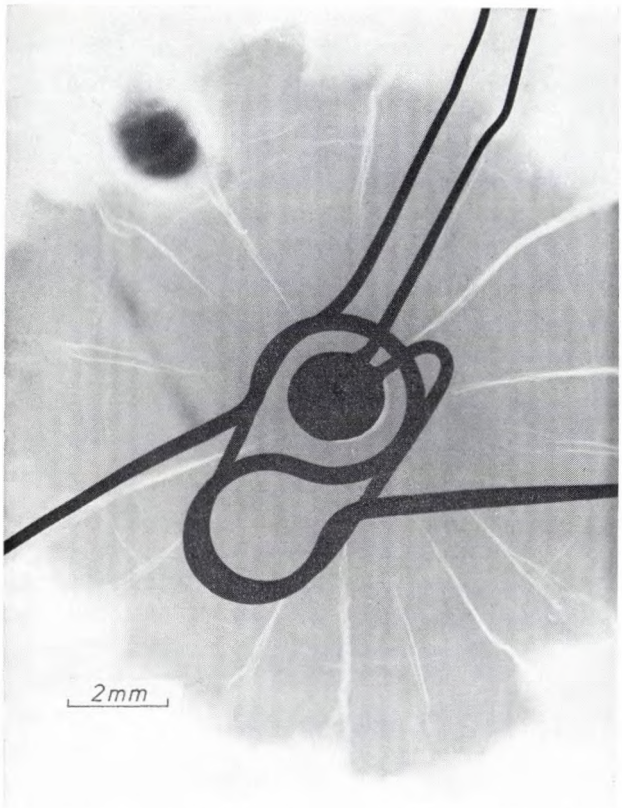


Fig. 11. X-ray photograph of the centre of the arrangement shown in fig. 10 after completion of an experiment. The cracks that form in the pyrophyllite after the pressure has been removed are clearly visible. The pressure was highest at the centre of the material, and the pyrophyllite has therefore been subject to the greatest compaction there; the photograph is darker in this region than at the edges.

of measurements by a number of investigators. It constitutes a calibration point, together with a number of other agreed phase transitions. This allows the work of different investigators to be compared and will permit later corrections to be applied when these phase transitions are measured more accurately in the future. Above 25 kbar the phase transitions used for calibration have been determined by correcting as accurately as possible for friction. For the phase transition taken as the calibration point for the highest pressure (77 kbar, also in Bi) this gives an accuracy of $\pm$ 3 kbar [10]. The phase transitions are generally found from a change in the electrical resistance or in the volume.

A secondary standard is calibrated from the phase transitions mentioned above. At room temperature the resistance change in manganin wire is generally adopted, whereas at liquid-helium temperatures the change in temperature at which tin becomes superconducting is usually adopted as the secondary standard. The method with manganin wire cannot be used with a solid pressure medium because the pressure is not the same over the entire length of the wire. In such a case the force on the plunger is first calibrated from the phase transitions mentioned above, using an assembly of components in the pressure chamber that corresponds as closely as possible to that of the actual experiment. This a much less accurate method than measuring the pressure at the specimen by means of a manganin coil, as was done in the liquid-filled 'Teflon' capsule of fig. 7.

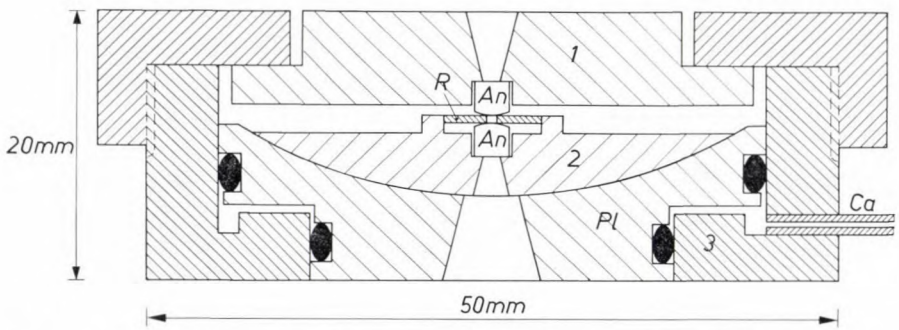


Fig. 12. Press with Bridgman anvils of diamond. The specimen is observed with a microscope through the diamonds *An*. Between the diamonds there is an 'Inconel' ring *R* with a hole of diameter 0.2 mm containing the liquid pressure medium and the specimen. The pressure surfaces of the diamonds are located accurately one above the other and parallel with the aid of parts *1* and *2*. The force required for this operation is provided by a gas under pressure introduced into the space between the plunger *Pl* and part *3* via the capillary tube *Ca*. Pressures of 500 kbar can be obtained with an experimental arrangement of this kind.

and the vessel. Although pressure can in fact be measured accurately in this way (to within 60 bars), special arrangements are required.

Such an arrangement has been used by the National Bureau of Standards in determining the phase transition in Bi at 25.5 kbar accurately [9]. In 1968 this transition was established at 25.50 $\pm$ 0.06 kbar at 25 °C. This value is the weighted mean of the results

When Bridgman anvils are used, it is completely impossible to derive the pressure from the force, since the pressure distribution over the anvil varies with the force. In such experiments calibrations are based on

[8] G. J. Piermarini and S. Block, Rev. sci. Instr. **46**, 973, 1975.
[9] P. L. M. Heydemann, J. appl. Phys. **38**, 2640, 1967.
[10] E. C. Lloyd (ed.), Accurate characterization of the high-pressure environment, Nat. Bur. Stand. spec. Publn No. 326, 1971.

the decrease in the lattice constant of NaCl (found by X-ray diffraction), which is correlated with the theoretically determined equation of state [11]. The calibration points are the pressures at which phase transitions occur in Ba, Pb, Fe, etc., which can easily be established from resistance measurements. The pressures are determined by simultaneously measuring the lattice constant of NaCl.

In optical work (vessels with windows, diamond anvils) the pressure can be determined accurately from the shift in the R_1 fluorescence line of ruby at 694.2 nm as a result of pressure: comparisons with the NaCl standard have shown that the shift is linear up to at least 290 kbar (+0.036 nm/kbar) [12]. At 290 kbar there is a phase change in NaCl. The pressure scale would therefore seem to have been established reasonably correctly up to 300 kbar.

[11] D. L. Decker, J. appl. Phys. **42**, 3239, 1971.
[12] J. D. Barnett, S. Block and G. J. Piermarini, Rev. sci. Instr. **44**, 1, 1973.
[13] A. S. Balchan and H. G. Drickamer, Rev. sci. Instr. **32**, 308, 1961.
[14] H. G. Drickamer, Rev. sci. Instr. **41**, 1667, 1970.

The nature of the developments in pressure-measurement techniques over the last 15 years can be illustrated most clearly by a few figures. In 1961, the scale that was then thought to apply [13] indicated that the semiconductor-metal transition of Si occurred at 195 kbar; in 1970 the best estimate was 150 kbar [14], and in 1975 the ruby method described above gave a value of 125 ± 5 kbar [8]. An even more striking example is provided by GaP, for which the 1970 scale [14] indicated a transition at about 500 kbar, whereas in 1975 Piermarini and Block found a pressure value of 220 ± 10 kbar.

Summary. In many areas of solid-state research high pressures are used to investigate the effect of change in volume on various properties of materials. The apparatus required for the measurements can be classified under three types. In order of increasing pressure these are the simple pressure vessel, the compound pressure vessel and the Bridgman anvils. Measurements with these different types of equipment can be made in a straightforward way and under well defined conditions up to pressures of 200 kbar in a temperature range from 2 to 1500 K. The three types are described in this article. Attention is also given to various special arrangements that are necessary for making measurements in a variety of fields.

Broadband circulators for VHF and UHF

G. Schiefer

Circulators and isolators are important components in microwave technology. Their essential feature is that they are non-reciprocal, so that they conduct energy in one direction only. The application of this property introduces considerable simplifications in the layout of radar and telecommunication equipment. In practice a large bandwidth is often required. Fundamental research at the Philips laboratories in Aachen has shown that circulators with lumped elements for VHF and UHF can be made with bandwidths up to an octave. It has also been demonstrated that the characteristics of broadband circulators can be kept constant in a wide temperature range, even though the saturation magnetization of the polycrystalline yttrium iron garnet used is highly temperature-dependent. A circulator has been developed that will operate from 225 to 400 MHz and in a temperature range from −20 °C to +80 °C.

Introduction

Circulators are non-reciprocal elements used in telecommunication engineering. Their characteristic feature is that they conduct energy in one direction only — in the usual three-port type, for example, from port *1* to port *2*, from port *2* to port *3* and from port *3* to port *1*; there is no energy transport in the opposite direction. They are used wherever different parts of a system have to be decoupled from each other, e.g. where a transmitter and a receiver use the same antenna or where the operation of an oscillator or amplifier must not be affected by signals reflected from the load.

Circulators utilize ferrimagnetic resonance in a magnetized ferrite disc. In addition to three-port junction circulators with waveguides or transmission lines arranged in the shape of a Y, circulators with lumped elements — i.e. inductors — have acquired a place for themselves in the frequency band below 1000 MHz [1].

In this frequency range the commercially available circulators have a relative bandwidth of about 10%, or possibly up to 20% if a small temperature range of operation is acceptable. For use at a fixed frequency such bandwidths are of course large enough. More recently, however, a demand has arisen for larger bandwidths, mainly for three reasons:

— broadband models permit a substantial reduction in the number of types;

— for military applications especially, increasing use is made of systems with continuously changing frequencies; the circulator cannot of course be changed every time the system is switched to a new frequency;

— the overcrowding of all frequency channels is making the problem of parasitic modulation more and more acute. This occurs where a transmitter receives external signals through its own antenna and then transmits them again cross-modulated. The only effective remedy is to introduce between antenna and transmitter a circulator or isolator that blocks the return path in the widest possible frequency band.

In principle the lumped-element circulator, which can be regarded as a non-reciprocal transformer with three windings, has a larger bandwidth than the Y circulator, which may be regarded as part of a waveguide or transmission line with the appropriate frequency-dependence. What is more, the lumped-element circulator can be made much smaller, since its dimensions have no relation to the wavelength as they have in the Y circulator. This article will therefore be concerned only with the lumped-element circulator, and will describe the measures that have enabled us to increase the bandwidth to nearly an octave. The prop-

Dr-Ing. G. Schiefer is with Philips GmbH Forschungslaboratorium Aachen, Aachen, West Germany.

[1] The operation and construction of circulators have been described in many publications. Examples are:
H. Bosma, On stripline Y circulation at UHF, IEEE Trans. MTT-12, 61-72, 1964;.
Y. Konishi, Lumped element Y circulator, IEEE Trans. MTT-13, 852-864, 1965;
H. Bex, Über konzentrierte Zirkulatoren, Thesis, Aachen 1969.
See also M. Lemke and W. Schilz, Microwave integrated circuits on a ferrite substrate, Philips tech. Rev. 32, 315-321, 1971.

erties of the ferrite used are highly temperature-dependent; with the right temperature compensation, however, the bandwidth is maintained over a relatively wide temperature range of operation.

Bandwidth of the lumped-element circulator

Fig. 1 is a schematic diagram of a symmetrical three-port lumped-element circulator. Three inductors consisting of one or more turns are mounted at angles of 120° on a disc of high-frequency ferrite (YIG). The coupling between the inductors is non-reciprocal when the disc is magnetically polarized in the axial direction by an external magnetic field. The behaviour has rotational symmetry, the sense of rotation depending on the direction of the polarization.

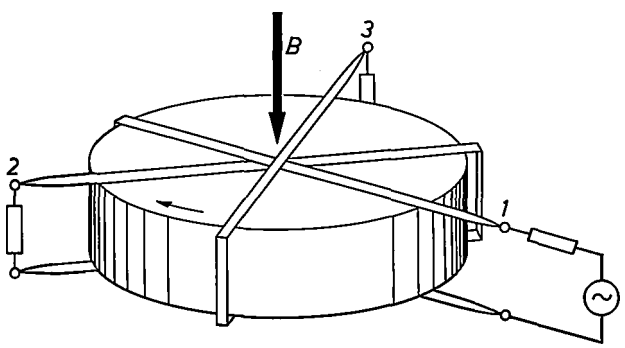


Fig. 1. Schematic representation of a lumped-element circulator. Three inductors with one or more turns are mounted at angles of 120° on a disc of high-frequency ferrite (YIG). Axial magnetization of the disc by means of an external magnetic field (flux density B) makes the coupling between the inductors non-reciprocal. When port *2* is connected to a matched (non-reflecting) load, the energy conducted by a generator to port *1* leaves the inductor only at port *2*, while port *3* is isolated.

The desired effect, which is to isolate a port (port *3* in this example), can only be achieved provided the output port (*2*) has a matched (i.e. non-reflecting) termination [2]. Otherwise the reflected energy would be conducted to port *3* and, on reflection there, would return to port *1*.

The bandwidth of a circulator is therefore essentially the frequency band over which the circulator ports can be matched sufficiently accurately to the external circuit. In practice the external circuit is generally equivalent to a resistance of 50 Ω. The bandwidth is then defined as the frequency band over which the circulator matched to 50 Ω gives no more than 10% reflection. This corresponds to a 20-dB reverse loss for the isolated port [3].

The impedance of the circulator must therefore be matched to 50 Ω within the widest possible frequency range. A three-port can in general be described with a matrix of nine impedances $Z_{11} \ldots Z_{33}$:

$$\begin{aligned} V_1 &= Z_{11}I_1 + Z_{12}I_2 + Z_{13}I_3, \\ V_2 &= Z_{21}I_1 + Z_{22}I_2 + Z_{23}I_3, \\ V_3 &= Z_{31}I_1 + Z_{32}I_2 + Z_{33}I_3, \end{aligned} \tag{1}$$

where V and I are the input voltages and currents at the three ports. For the circulator there is rotational symmetry, so that this matrix may be simplified to

$$\begin{pmatrix} Z_{11} & Z_{12} & Z_{13} \\ Z_{13} & Z_{11} & Z_{12} \\ Z_{12} & Z_{13} & Z_{11} \end{pmatrix}. \tag{2}$$

For an ideal lossless circulator the elements of this matrix are:

$$Z_{11} = \omega L_0 \mathrm{j}\mu, \tag{3}$$

$$Z_{12} = \tfrac{1}{2}\omega L_0(\kappa\sqrt{3} - \mathrm{j}\mu), \tag{4}$$

$$Z_{13} = -\tfrac{1}{2}\omega L_0(\kappa\sqrt{3} + \mathrm{j}\mu), \tag{5}$$

where ω is the angular frequency, L_0 the air inductance of a single inductor; μ and κ are the elements of a tensor with which the permeability of the ferrite is described. For the case $V_3 = 0, I_3 = 0$ (port *3* isolated) it then follows that the impedance of the circulator is:

$$Z_c = \tfrac{3}{2}\omega L_0 \frac{\kappa^2 - \mu^2}{3\kappa^2 + \mu^2}(\kappa\sqrt{3} + \mathrm{j}\mu). \tag{6}$$

For the case of no losses we may write [4]:

$$\mu = 1 + \frac{\gamma^2 H_i M_s}{\gamma^2 H_i^2 - \omega^2}, \tag{7}$$

$$\kappa = -\frac{\gamma M_s \omega}{\gamma^2 H_i^2 - \omega^2}, \tag{8}$$

where H_i is the internal field in the ferrite disc, M_s is the saturation magnetization of the ferrite and $\gamma = 2\pi \times 3.5$ MHz/(A/cm) is the gyromagnetic ratio.

The admittance of the circulator is thus

$$Y_c = \frac{1}{Z_c} = \frac{2}{\sqrt{3}}\frac{1}{L_0}\frac{\gamma M_s}{\gamma^2(H_i + M_s)^2 - \omega^2} - \mathrm{j}\tfrac{2}{3}\frac{1}{\omega L_0}\frac{\gamma^2 H_i(H_i + M_s) - \omega^2}{\gamma^2(H_i + M_s)^2 - \omega^2}. \tag{9}$$

Anticipating the final result, we see the importance of the case $\omega \approx \gamma H_i \ll \gamma M_s$. For this case we find:

$$Y_c \approx \frac{2}{L_0 \gamma M_s \sqrt{3}} - \mathrm{j}\frac{1}{\omega}\frac{2H_i}{3L_0 M_s}. \tag{10}$$

To a first approximation the circulator admittance can clearly be described as the parallel configuration of a

(frequency-independent) resistance with an inductance (*fig. 2*).

For a circuit of this type the width $\Delta\omega$ of a frequency band around a central angular frequency ω_0 can be indicated within which the circulator reaches an isolation value of 20 dB [3]. In the case of a first-order matching with a capacitor connected in parallel (possibly tapped down to give impedance transformation) we may write with the same approximation that led to eq. (10):

$$\frac{(\Delta\omega)_1}{\omega_0} \leqslant 0.2\frac{\omega_0\sqrt{3}}{\gamma H_i}. \tag{11}$$

For second-order matching with at least three reactances we have:

$$\frac{(\Delta\omega)_2}{\omega_0} \leqslant 0.7\frac{\omega_0\sqrt{3}}{\gamma H_i}. \tag{12}$$

Higher-order matching with more reactances is not very useful in practice, since the bandwidth increases only very slightly whereas adjustment becomes difficult and the losses too large.

Effect of the air gap

In practice the bandwidths given by equations (11) and (12) cannot be reached since the idealized lossless

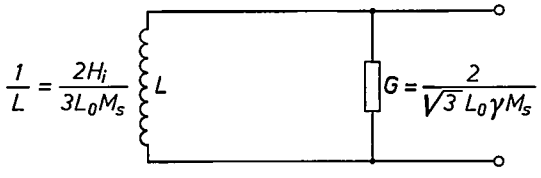


Fig. 2. Equivalent circuit of the admittance of the lumped-element circulator. This admittance corresponds to that of a resistor of conductance G in parallel with an inductance L.

circulator cannot be made in practice. Magnetic leakage occurs, since the fields of the three inductors are not enclosed completely in the ferrite but have to cross air gaps because of the geometry.

The effect of this can be investigated, at least qualitatively, by describing the action of the circulator in terms of two fields rotating in opposite directions, each of which encounters a different permeability in the ferrite disc [5]:

$$\mu_r = \mu - \kappa = 1 + \frac{\gamma M_s}{\gamma H_i - \omega}, \tag{13}$$

and

$$\mu_1 = \mu + \kappa = 1 + \frac{\gamma M_s}{\gamma H_i + \omega}. \tag{14}$$

As in the case of ferrite inductors with an air gap, the demagnetization is expressed by introducing an effective permeability; for reasons of symmetry the same demagnetizing factor N holds for μ_r and μ_1:

$$\mu_{1,r,\mathrm{eff}} = \frac{\mu_{1,r}}{1 + (\mu_{1,r} - 1)N}. \tag{15}$$

In

$$\mu_{\mathrm{eff}} = \frac{\mu_{1,\mathrm{eff}} + \mu_{r,\mathrm{eff}}}{2} \tag{16}$$

and

$$\kappa_{\mathrm{eff}} = \frac{\mu_{1,\mathrm{eff}} - \mu_{r,\mathrm{eff}}}{2} \tag{17}$$

we then have effective values which can be introduced into equations (3) to (6) instead of μ and κ. Using the approximations adopted for eq. (10) we then find instead of eqs. (11) and (12):

$$\frac{(\Delta\omega)_1}{\omega_0} \leqslant 0.2\frac{(1-N)\omega_0\sqrt{3}}{\gamma(H_i + NM_s)}, \tag{18}$$

$$\frac{(\Delta\omega)_2}{\omega_0} \leqslant 0.7\frac{(1-N)\omega_0\sqrt{3}}{\gamma(H_i + NM_s)}. \tag{19}$$

The introduction of the demagnetizing factor N thus leads to smaller results for the calculated bandwidth.

The geometry of the coupled inductors

Equations (18) and (19) are admittedly rough approximations, but they clearly show how important it is when designing broadband circulators to make the internal field H_i and the demagnetizing factor N as small as possible.

A lower limit is set to the field-strength H_i in the ferrite mainly by the occurrence of losses when the field-strength is too low. These occur when the entire volume of the ferrite disc is no longer magnetically saturated. The result is a sudden increase in the forward loss of the circulator.

To achieve full saturation of the ferrite disc even for weak fields, two main conditions have to be met: the ferrite must have a low coercivity and a sharp bend in its hysteresis loop at saturation, and the geometry of the entire magnetic circuit must be optimized in such a way as to obtain the same induced field-strength throughout the disc.

---

[2] This simplified description only holds exactly for a lossless circulator. It is however sufficiently accurate for the present purposes for a practical low-loss circulator.

[3] E. Schwartz, Broadband matching of resonant circuits and circulators, IEEE Trans. MTT-16, 158-165, 1968.

[4] B. Lax and K. J. Button, Microwave ferrites and ferrimagnetics, McGraw-Hill, New York 1962.

[5] H. Bex and E. Schwartz, Wirkungsweise konzentrierter Zirkulatoren, Frequenz 24, 288-293, 1970.

The first requirement can be met particularly well with polycrystalline substituted YIG materials: these have been developed at Philips Forschungslaboratorium Hamburg [6]. In practice the magnetic circuit can only be optimized experimentally; the following conditions are important. Firstly, the YIG discs must be very thin (diameter/thickness > 20). Secondly, the magnetic circuit must have soft-iron pole pieces whose diameter is at least 15% greater than that of the ferrite discs. And thirdly, pole pieces and ferrite discs must be accurately plane-parallel.

To keep the demagnetizing factor N small, close coupling is necessary between the three inductors and between the inductors and the ferrite. The design shown in *fig. 3*, developed in our laboratories, has proved particularly suitable for this purpose; in each inductor the forward path for the current consists of four or more parallel conductors situated between two YIG discs, and the return path for all three inductors together consists of two thin metal plates outside the YIG discs. To obtain better symmetry the conductors are threaded through one another. The optimum width and spacing of the separate conductors and their distance from the ferrite are sometimes critical and have to be determined experimentally.

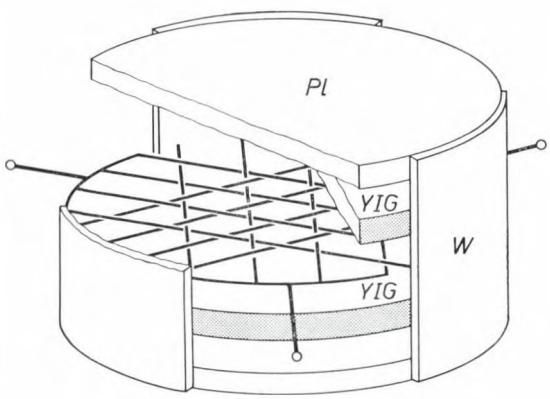


Fig. 3. Arrangement of the inductors in the broadband circulator. The forward path for the current in each inductor consists of four parallel conductors; the common return path is formed by the vertical walls *W* and the two plates *Pl* of sheet metal. The interwoven inner conductors are located between two ferrite discs *YIG*.

With the design described here practical values for $\gamma(H_i + NM_s)$ as small as $2\pi \times 300$ MHz have been achieved, with N-values smaller than 0.05. The bandwidths obtainable with these values can easily be estimated from equations (18) and (19). At centre frequencies above 300 MHz, second-order matching gives bandwidths of the order of one octave.

Matching networks

As already described, the equivalent impedance of the circulator is well represented by a resistance and an inductance in parallel (fig. 2). For matching to the standard 50-ohm lines an impedance-transforming network has to be introduced in the circuit for each circulator port. The circuit configuration depends on whether the conductance G of the resistance is greater than, equal to or smaller than $1/(50\,\Omega)$ in first-order matching (or $1/(60\,\Omega)$ in second-order matching). Typical matching networks for such cases are shown in *fig. 4*. The method of calculating the separate network elements is well known and will not be dealt with here.

Forward loss

The forward loss of the circulator is attributable to losses in the ferrite, in the coupling inductors and in the matching networks. The maximum permissible forward loss for circulators for small and medium powers is generally 1 dB. For high powers a lower forward loss is usually required, since the efficiency and the admissible heating are then important.

A fraction α_f of the forward loss is due to the ferrite, even when there is complete saturation; it depends mainly on the relaxation losses and is manifested in the linewidth ΔH of the ferrimagnetic resonance. A useful approximation for this part of the forward loss is [7]:

$$\alpha_f \approx 2.5 \frac{\gamma \Delta H}{\omega} \text{ dB}. \qquad (20)$$

In broadband circulators the lowest frequency is of the greatest importance. The conclusion arrived at is that, particularly in the VHF band, the only materials that can be used are those with a linewidth of less than 10 A/cm. This condition is met only by the doped polycrystalline YIG materials mentioned above or, in extreme cases, by YIG single crystals.

In practice, however, it is also necessary to take into account the additional losses that occur at low field-strengths, since complete saturation of the YIG discs is not feasible at the field-strengths desirable here. The value mentioned earlier of

$$\gamma(H_i + NM_s) \approx 2\pi \times 300 \text{ MHz}$$

was obtained with a magnetization at which the total forward loss attributable to losses in the ferrite was less than 1 dB.

Compared with this the forward loss due to the coupling inductors and the matching networks is in most cases negligible. This is readily understood when we remember that the Q (quality factor) of the admittance of broadband circulators is about 1, whereas the Q of the components in the circuits is usually higher

than 100. What is important, however, is always to choose the circuit in which the impedance transformation takes place by the shortest route [8].

Design of the magnetic circuit

The saturation magnetization of high-frequency ferrites decreases considerably with rising temperature. For YIG material the curve in the temperature range of technical interest is well represented by a linear decrease in the saturation magnetization: $\mu_0 dM_s/dT \approx -7\times10^{-4}$ tesla/°C ($= -7$ gauss/°C). This implies, for example, a change in the saturation flux density of 0.20 tesla at −20 °C to 0.13 tesla at 80 °C.

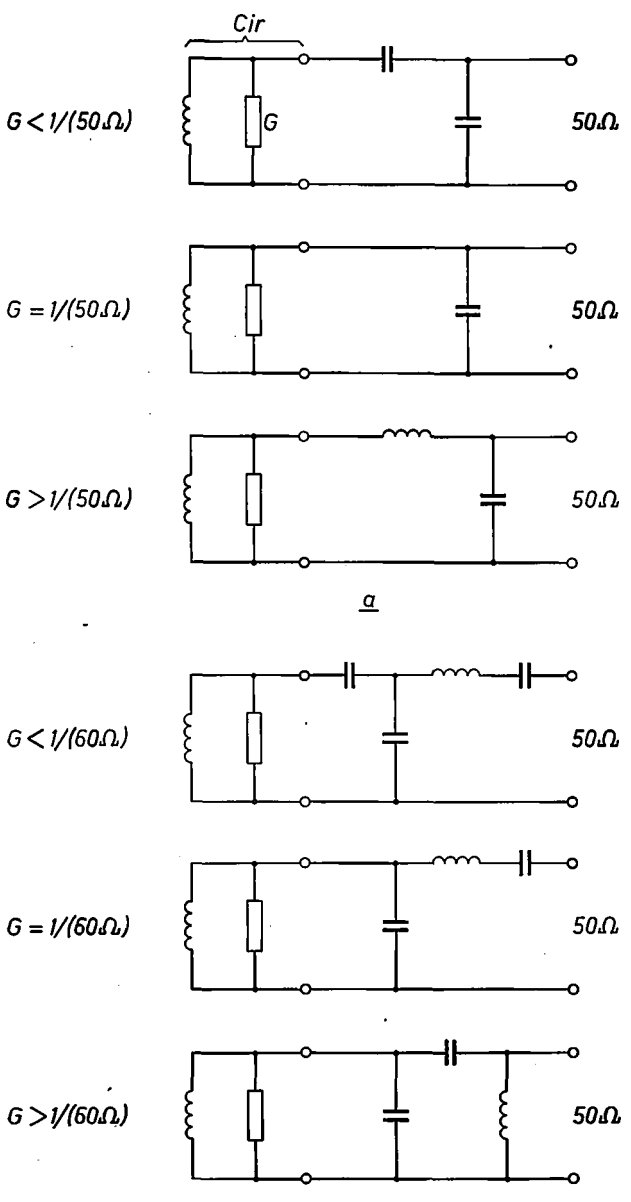


Fig. 4. Networks for matching the circulator *Cir* to a 50-Ω line. The networks differ for different values of G, the real part of the admittance of the circulator (fig. 2). *a*) First-order matching (for one frequency). *b*) Second-order matching.

From eq. (10) it appears that the real part of the admittance of the circulator is inversely proportional to M_s. Matching is therefore only possible by means of a network that has a corresponding temperature dependence, as described in the next section. The imaginary part on the other hand is dependent on H_i/M_s. This ratio can be kept constant if the magnetic circuit is designed in such a way that H_i has the same temperature dependence as M_s. We then have:

$$M_s(T) = M_s(T_0) + \frac{dM_s}{dT}(T-T_0), \qquad (21)$$

$$H_i(T) = H_i(T_0) + \frac{H_i(T_0)}{M_s(T_0)}\frac{dM_s}{dT}(T-T_0). \qquad (22)$$

From the ideal hysteresis curve of the ferrite we then derive the flux density $B_f(T)$ that the magnetic circuit should provide between its poles as a function of the temperature:

$$B_f(T) = B_f(T_0) + \frac{dB_f}{dT}(T-T_0) = \mu_0\{H_i(T) + M_s(T)\}$$

$$= \mu_0\left[H_i(T_0) + M_s(T_0) + \left\{1 + \frac{H_i(T_0)}{M_s(T_0)}\right\}\frac{dM_s}{dT}(T-T_0)\right]. \qquad (23)$$

Normally the temperature dependence of materials for permanent magnets is less than that of high-frequency ferrites. The desired large variation in $B_f(T)$ with temperature can therefore only be achieved by introducing into the magnetic circuit by way of compensation a component whose temperature dependence is substantially greater than that of the ferrite. The densely sintered ferroxcube 3E1 has proved very suitable for this purpose; its saturation magnetization decreases as $\mu_0 dM_s/dT = -28.5\times10^{-4}$ tesla/°C.

The calculation of the dimensions of the temperature-compensated magnetic circuit will now be presented briefly. The total circuit consists of a series arrangement of pairs of discs with an overall thickness l of high-frequency ferrite (subscript f), compensation ferrite (subscript c), magnet material (subscript m) and of air gaps (subscript 0), together with an iron yoke that can be disregarded here.

[6] G. Winkler, P. Hansen and P. Holst, Variation of the magnetic material parameters and lattice constants of polycrystalline yttrium-iron garnet by incorporation of nonmagnetic ions, Philips Res. Repts. 27, 151-171, 1972.
[7] See the thesis by H. Bex, note [1].
[8] More details will be found in two articles by E. Schwartz: Zur Theorie der Anpassung mit zwei Reaktanzen, AEÜ 23, 169-176, 1969, and Die Bandbreite von Anpassungsvierpolen mit zwei Reaktanzen, AEÜ 24, 179-186, 1970.

The magnetic fields in the various discs are given by:

$$H_{\mathrm{im}} l_{\mathrm{m}} = H_0 l_0 + H_{\mathrm{if}} l_{\mathrm{f}} + H_{\mathrm{ic}} l_{\mathrm{c}}, \tag{24}$$

and the magnet fluxes, when A represents the cross-sectional area in each case, by:

$$B_{\mathrm{m}} A_{\mathrm{m}} = B_0 A_0 = B_{\mathrm{f}} A_{\mathrm{f}} = B_{\mathrm{c}} A_{\mathrm{c}}. \tag{25}$$

The values of B are derived from ideal hysteresis loops:

$$\begin{aligned} B_{\mathrm{m}} &= B_{\mathrm{r}} - \mu_0 H_{\mathrm{im}} \text{ (for ferroxdure 330)}, \\ B_0 &= \mu_0 H_0, \\ B_{\mathrm{f}} &= \mu_0 (H_{\mathrm{if}} + M_{\mathrm{sf}}), \\ B_{\mathrm{c}} &= \mu_0 (H_{\mathrm{ic}} + M_{\mathrm{sc}}). \end{aligned} \tag{26}$$

B_{r} is the remanence of the magnet material (in this case ferroxdure 330). The ratios of the cross-sectional areas follow from the selected points on the characteristics. B_{m} can best be selected such that the operating point at the highest remanence value B_{r}, i.e. at the lowest desired temperature T_{min}, only just lies on the straight part of the characteristic; see *fig. 5*. From this we find

$$\sigma = \frac{A_{\mathrm{f}}}{A_{\mathrm{m}}} = \frac{B_{\mathrm{m}}}{B_{\mathrm{f}}} \quad (T = T_{\mathrm{min}}). \tag{27}$$

Similarly, the cross-sectional area of the compensating material must be selected in such a way that the material is still fully saturated at the lowest temperature:

$$\tau = \frac{A_{\mathrm{f}}}{A_{\mathrm{c}}} = \frac{B_{\mathrm{c}}}{B_{\mathrm{f}}} \quad (T = T_{\mathrm{min}}). \tag{28}$$

The areas of the ferrite and the air gap can be regarded to a first approximation as identical:

$$A_0 = A_{\mathrm{f}}. \tag{29}$$

Substitution in eq. (24) gives:

$$(B_{\mathrm{r}} - \sigma B_{\mathrm{f}}) l_{\mathrm{m}} = B_{\mathrm{f}} l_0 + \mu_0 H_{\mathrm{if}} l_{\mathrm{f}} + (\tau B_{\mathrm{f}} - \mu_0 M_{\mathrm{sc}}) l_{\mathrm{c}}. \tag{30}$$

This equation holds for a particular temperature T with the appropriate values of B_{r}, B_{f}, H_{f} and M_{sc}. If the desired temperature compensation is to be obtained, however, this equation must also hold for changes in these quantities with temperature. Using equations (21), (22) and (23) and also $B_{\mathrm{r}}(T) = B_{\mathrm{r}}(T_0) + (\mathrm{d}B_{\mathrm{r}}/\mathrm{d}T)(T - T_0)$ and $M_{\mathrm{sc}}(T) = M_{\mathrm{sc}}(T_0) + (\mathrm{d}M_{\mathrm{sc}}/\mathrm{d}T)(T - T_0)$, then

$$\begin{aligned} &\left[\frac{\mathrm{d}B_{\mathrm{r}}}{\mathrm{d}T} - \sigma\mu_0 \frac{\mathrm{d}M_{\mathrm{sf}}}{\mathrm{d}T}\left\{1 + \frac{H_{\mathrm{if}}(T_0)}{M_{\mathrm{sf}}(T_0)}\right\}\right] l_{\mathrm{m}} = \mu_0 \frac{\mathrm{d}M_{\mathrm{sf}}}{\mathrm{d}T}\left\{1 + \frac{H_{\mathrm{if}}(T_0)}{M_{\mathrm{sf}}(T_0)}\right\} l_0 \\ &+ \mu_0 \frac{H_{\mathrm{if}}(T_0)}{M_{\mathrm{sf}}(T_0)} \frac{\mathrm{d}M_{\mathrm{sf}}}{\mathrm{d}T} l_{\mathrm{f}} + \\ &+ \left[\tau\mu_0 \frac{\mathrm{d}M_{\mathrm{sf}}}{\mathrm{d}T}\left\{1 + \frac{H_{\mathrm{if}}(T_0)}{M_{\mathrm{sf}}(T_0)}\right\} - \mu_0 \frac{\mathrm{d}M_{\mathrm{sc}}}{\mathrm{d}T}\right] l_{\mathrm{c}}. \end{aligned} \tag{31}$$

Finally, from the two equations (30) and (31) we can calculate the length l_{m} of the permanent magnet and the length l_{c} of the compensating material.

In practice, the temperature compensation is always bound to be a compromise. Allowing a change in H_{i} implies, as (18) and (19) show, that the largest bandwidth is only achieved at the smallest H_{i}, i.e. at the highest temperature. Provided the values of the magnetic circuit are suitably chosen, it is possible to keep H_{i} almost constant during changes of temperature. This means, however, that the imaginary part of the circulator admittance then becomes temperature-dependent, and experience has shown that this makes

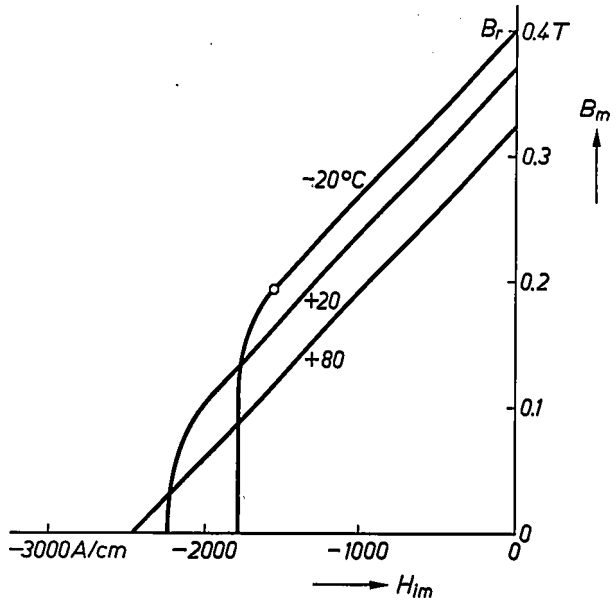


Fig. 5. Choice of the operating point (H_{im},B_{m}) on the characteristic of the permanent magnet. B_{m} can best be chosen such that the operating point at the highest remanence value B_{r}, i.e. at the lowest temperature, is only just on the straight part of the characteristic.

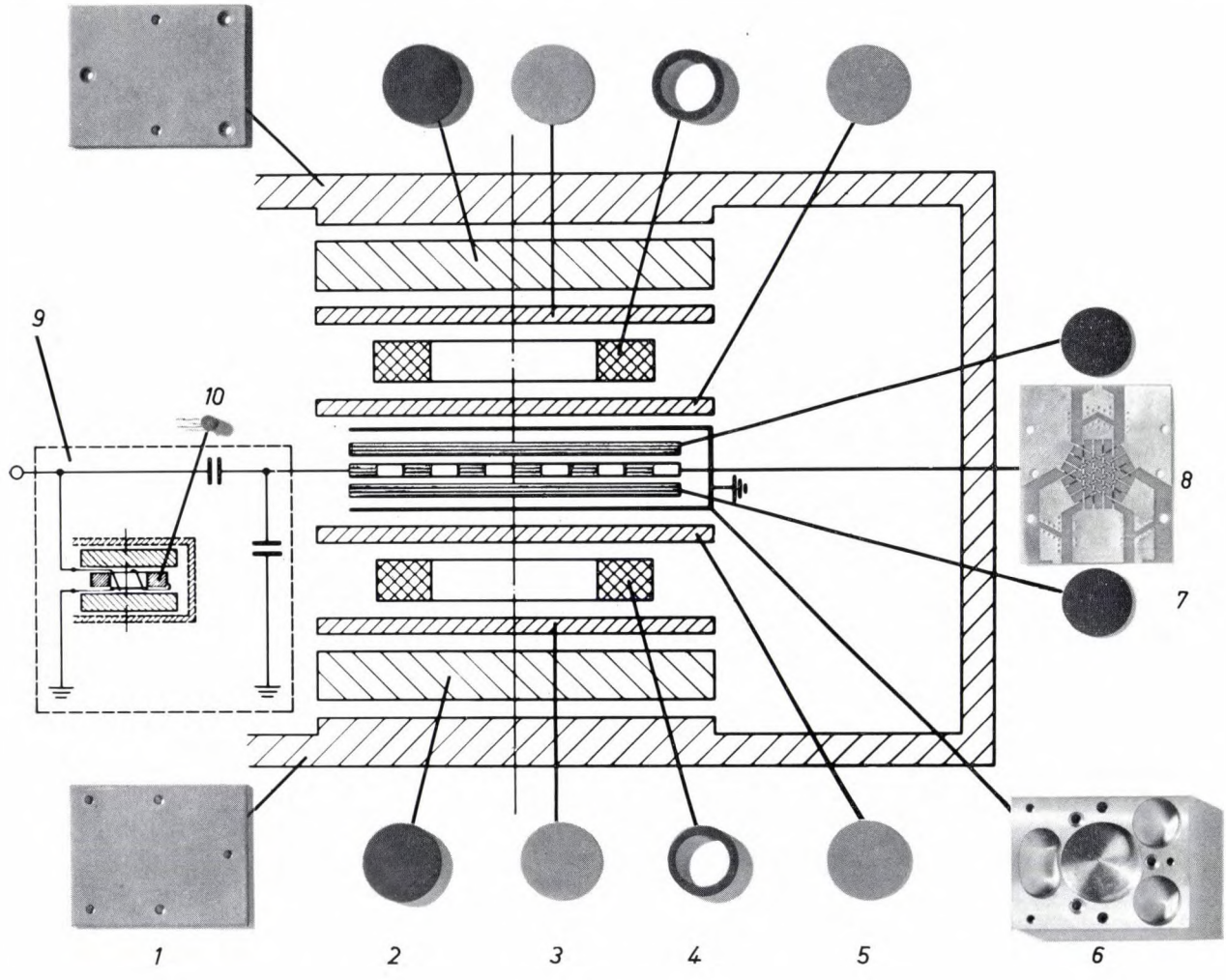
the compensation in the network very much more difficult to achieve. The optimum design usually lies between the two cases mentioned and the final details have to be determined experimentally. In any case, it must be borne in mind that the originally calculated bandwidth cannot be fully utilized in a wider temperature range.

The calculation of the magnetic circuit is incorrect if there is magnetic leakage. A good way of avoiding a leakage flux is to introduce the compensating material in the form of a ring between two sufficiently large soft-iron pole plates. An effective version of the complete magnetic circuit is shown in *fig. 6*.

Temperature compensation of the matching network

The change in the real part and possibly in the imaginary part of the circulator admittance with temperature can be compensated within certain limits by forming the matching networks given in fig. 4 from components with the 'correct' temperature coefficient. No general rules can be given here for calculating this temperature coefficient. It must be calculated separately for each separate case from measurements of the circulator admittance for at least three different temperatures.

Depending on the choice of the method of impedance transformation, components may be required with a relatively high positive or negative temperature coefficient, up to a few 10^{-3}/°C. Although capacitors with values up to about $\pm 1.5 \times 10^{-3}$/°C are commercially



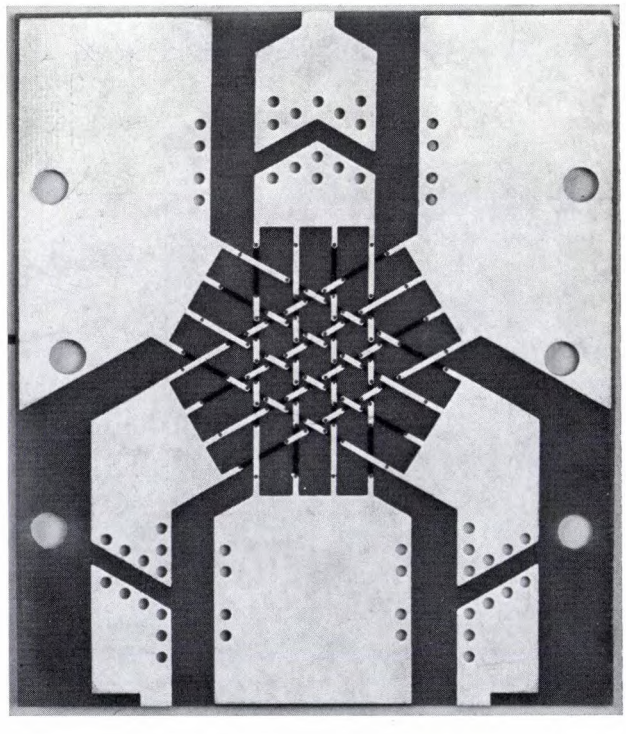


Fig. 6. Cross-section of the broadband circulator and magnified view of the inner conductors.

1 iron yoke
2 ferroxdure 330 magnet
3 pole piece
4 compensation ring of ferroxcube 3E1
5 pole piece
6 outer conductor
7 YIG disc
8 inner conductors
9 matching network
10 inductor with negative temperature coefficient

available, inductors generally have only a small positive temperature coefficient. The development of inductors with a greater temperature dependence will therefore now be briefly described.

Fig. 7 shows a cross-sectional view of a toroidal wire-wound ring of high-frequency ferrite placed

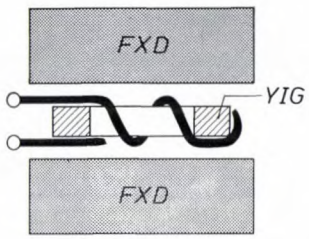


Fig. 7. Inductor with high negative temperature coefficient. A toroidal wire-wound ring of high-frequency ferrite *YIG* is contained between two permanent magnets *FXD* of ferroxdure.

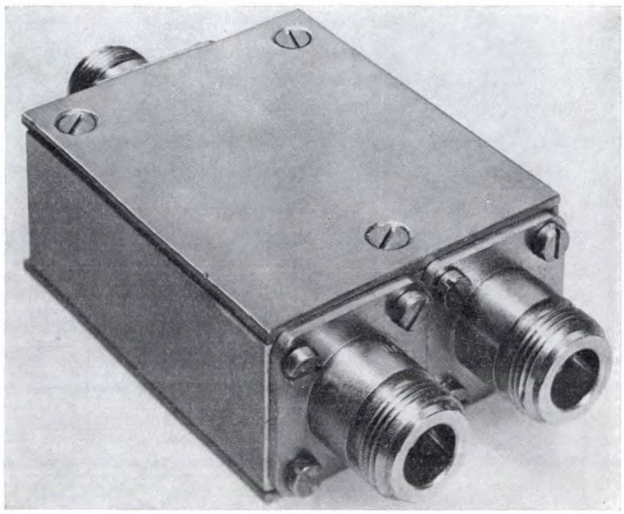


Fig. 8. The broadband circulator for the frequency band 225 MHz-400 MHz, which can be used in a temperature range from −20 °C to +80 °C.

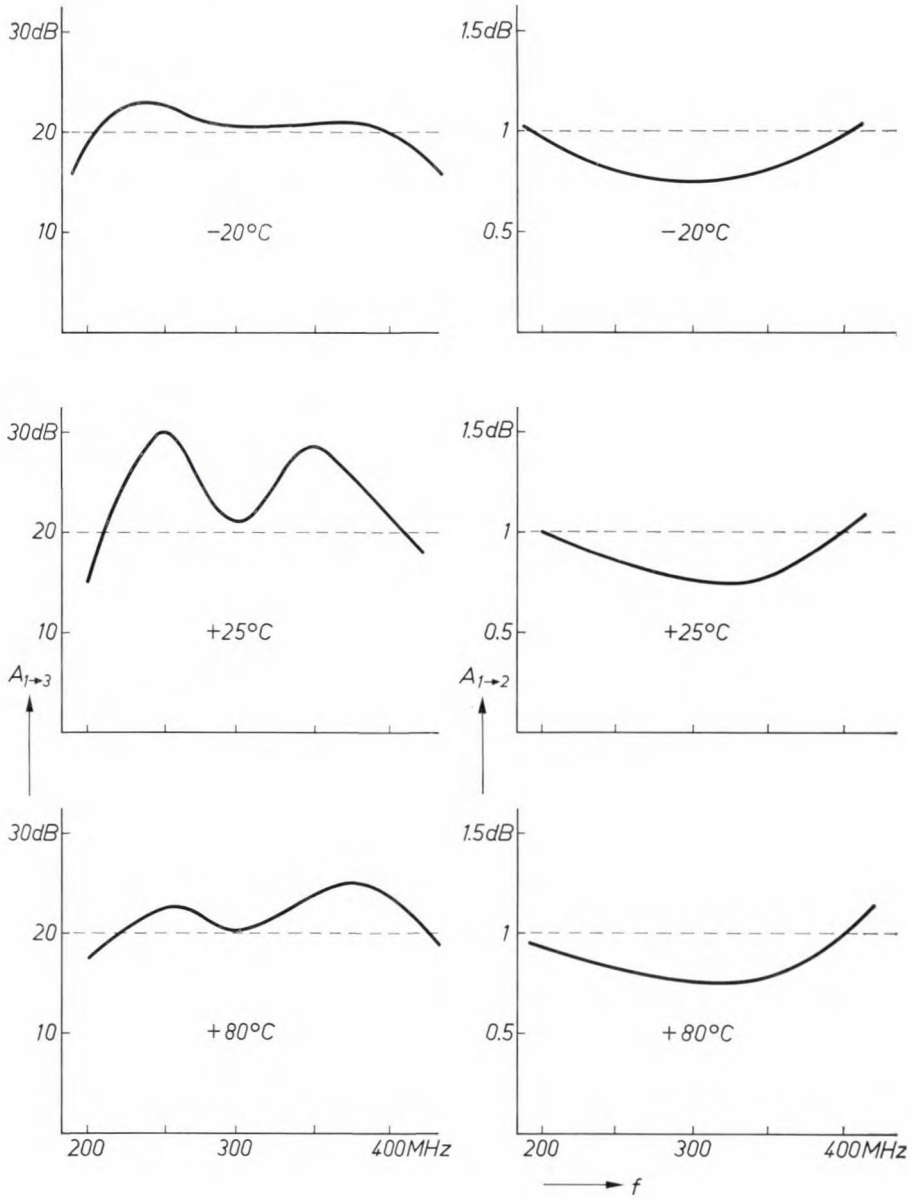


Fig. 9. Reverse loss $A_{1\to3}$ and forward loss $A_{1\to2}$ of the VHF/UHF circulator of fig. 8 as a function of frequency f for three different temperatures. The forward loss is in all cases less than 1 dB, the reverse loss greater than 20 dB.

coaxially between two non-conducting permanent magnets of ferroxdure. The inductance of the winding is proportional to the relative permeability μ_r of the ferrite ring. For sufficiently high magnetization by the magnetic flux of density B, this is given by:

$$\mu_r \approx 1 + \frac{M_s}{H_i} = 1 + \frac{\mu_0 M_s}{B - \mu_0 M_s}. \qquad (32)$$

Since, as described in the previous section, the decrease in the saturation magnetization M_s of the ferrite with temperature is much greater than that of the flux density B, the value of μ_r thus decreases with temperature. With a suitable choice of material and design inductors with a negative temperature coefficient up to a few 10^{-3}/°C can be produced in this way.

This design also allows positive temperature coefficients to be obtained by placing a compensating ferrite whose saturation magnetization is highly temperature-dependent between the magnets and the ferrite ring. The details can also be found in the previous section.

A circulator for the VHF band

Finally some data will be presented relating to a circulator developed along the lines described above for the professional VHF band of 225-400 MHz [9]. The complete circulator is shown in *fig. 8*; a cross-sectional view and the components can be seen in fig. 6. The interwoven inner conductors are photolithographically etched from a plastic substrate coated on both sides with metal. The alignment of the matching networks is fairly critical and at each port at least the series capacitance and the inductor with the negative temperature coefficient should be made variable. The values of the forward and reverse losses obtained with this model are given in *fig. 9* for three different temperatures. At all three temperatures the circulator meets the specifications of a forward loss less than 1 dB and a reverse loss greater than 20 dB in the complete band from 225 to 400 MHz.

The power transfer is limited by effects that occur in the ferrite. Above a certain critical field-strength spin waves are excited in the ferrite. This makes the transfer characteristics of the circulator nonlinear, so that intermodulation can arise. The power-handling capacity of the circulator is also limited by the heat developed; however, this can be kept under control by suitable provisions for heat dissipation The laboratory model described here can handle a power of about 50 watts.

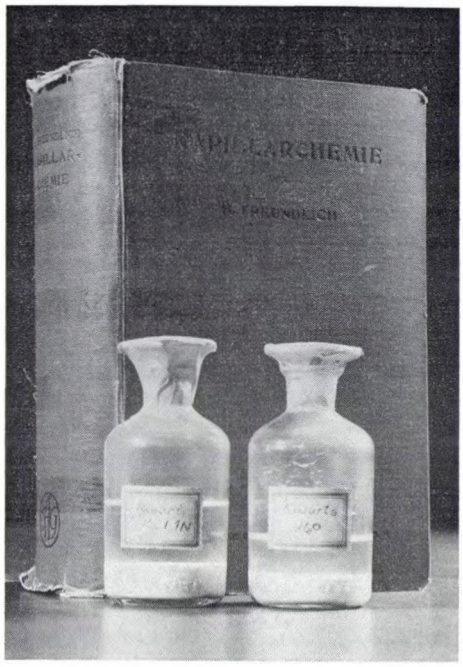
[9] In cooperation with Valvo, Hamburg; see F. Pötzl, VHF-Breitbandzirkulatoren in konzentrierter Bauweise, Valvo Berichte **19**, 19-28, 1974.

Summary. Starting from the description of a circulator with lumped elements procedures are derived for the development of a broadband VHF/UHF circulator that can be used within a wide range of temperatures. A circulator has been built for the frequency band from 225 to 400 MHz and a temperature range from −20 °C to +80 °C. The circulator consists of three interwoven conductors between two YIG discs. The bandwidth of such a circulator increases as the magnetizing field in the YIG discs decreases. In spite of the high temperature dependence of the saturation magnetization of YIG, this weak field is kept almost constant by including rings of densely sintered ferroxcube 3E1 in the magnetic circuit. The impedance at each port is transformed to 50 ohms by two capacitors with a positive temperature coefficient and a specially developed inductor with a high negative temperature coefficient.

Philips tech. Rev. **36**, 264-270, 1976, No. 9

Suspension technology

G. Frens, H. F. Huisman, J. K. Vondeling and K. M. van der Waarde



In 1909 H. Freundlich was the first to describe the effects observed in colloidal solutions and suspensions as 'capillary chemistry'. Since then suspension technology has undergone vigorous development, based on a growing understanding of the operation of capillary forces between particles and the ways in which they can be manipulated. In production processes of various kinds the surface tension and stability of suspensions are used as 'tools' for the application of layers of particles with a well defined structure.

The capillary forces in a suspension

A suspension is a collection of small particles floating around in a medium. Smoke is a suspension of soot in air, milk is a suspension of fat in water, and paint is a suspension of pigment particles in oil. Between the particles in a suspension interaction forces exist, known as capillary forces. These forces exert their action in situations where there are interfaces between materials. A familiar example of a capillary force is surface tension, which can draw up a column of liquid in a narrow tube when the liquid wets the wall of the tube.

In suspensions the surface areas are large in relation to the volumes, characteristically of the order of square metres per gram of material. What happens in such systems depends on capillary forces. The suspended particles may repel or attract one another; in the latter case they may form aggregates — the suspension 'flocculates' or curdles.

There are various types of capillary force. In one type the force is directly connected with the nature of the substance of which the suspended particles are made. This group includes the London-Van der Waals forces — or 'dispersion forces' — and also the forces that magnetic particles exert on each other. Around the particle there is a field, which extends a certain distance into the surrounding medium.

Another type of capillary force is the consequence of adsorption. At each interface there will be an accumulation of all kinds of charges, ions and 'surface-active' substances (surfactants). Their presence is also perceptible at some distance away from the interface: on interfaces there are electrical double layers and also layers of adsorbed molecules whose thickness may even exceed the dimensions of the suspension particles themselves.

A suspension is 'stable' when the particles have no tendency to form aggregates. Since there are always *attractive* capillary forces between the particles in a suspension, whether or not a stable suspension can be obtained depends on whether it is possible to bring repulsive forces into play.

Repulsive capillary forces between particles are usually of the second type; they are connected with adsorption. It is not difficult to see that the adsorption of substances on the surface of the particles can give rise to repulsion: the closer the two surfaces approach one another, the higher is the concentration of the

Dr G. Frens and Ing. J. K. Vondeling are with Philips Research Laboratories, Eindhoven; Dr H. F. Huisman is with the Magneetbandfabriek Oosterhout B.V., Oosterhout; Drs K. M. van der Waarde is with the Philips Lighting Division, Eindhoven.

adsorbed molecules or ions in the volume between the interfaces. The osmotic value of the medium between the particles is then greater than elsewhere in the medium, and an osmotic pressure exists that tries to drive the particles away from one another to a distance where the adsorption effects are no longer perceptible in the medium. Both the repulsion between two electrical double layers [1] — where the medium is enriched with adsorbed 'counter'-ions over a distance equal to the Debye length — and the repulsion between two layers of adsorbed macromolecules [2] ('entropic repulsion') can be explained in this way.

The various capillary forces differ not only in strength but also in the way in which they depend on the distance to the particle that exerts the force. What happens when forces of different kinds are acting simultaneously can be determined by investigating the way in which the free energy V of two particles depends on the distance between them [3]; see *fig. 1*. Repulsive forces provide a contribution (V_R) that increases as the particles approach one another, attractive forces a contribution (V_A) that decreases. The net effect of attraction and repulsion on the free energy is found by summation of V_R and V_A. If the sum is smaller for shorter than for longer distances, the particles tend to aggregate: their free energy V is lower when they are close together. In the situation illustrated in fig. 1, V does not vary monotonically with distance but has a maximum. There is then an energy barrier that tends to prevent the particles from coming together.

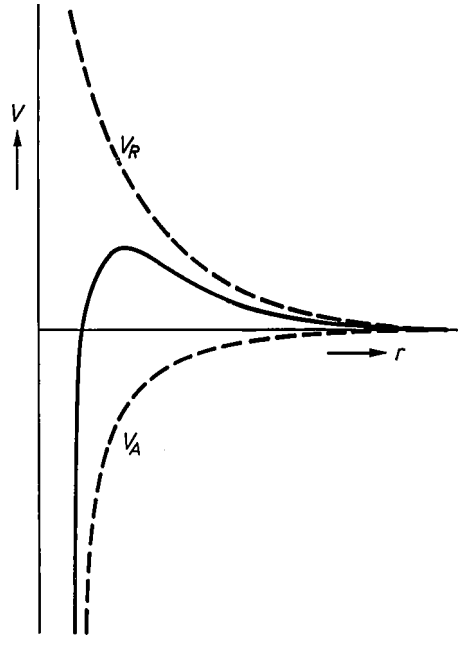


Fig. 1. Free energy V as a function of the distance r between two particles. The sum of the attraction and repulsion contributions ($V_A + V_R$) in the case shown here leads to an energy barrier. Only under this condition can a suspension be stable; if the repulsion is weaker, V decreases monotonically as the particles come closer to each other and the suspension is unstable.

As already indicated, it is only the *repulsive* forces that can be utilized to obtain a stable suspension. The magnitude and the range of the *attractive* forces between the particles of a given material are fixed; there is no way in which they can be varied. The required repulsive forces are obtained, as we have seen, from adsorption. Not only can these forces be deliberately generated, they can also be applied in the required magnitude; by controlling the adsorption it is possible to make a suspension stable or unstable, whichever is required. The suspension technologist makes wide use of this possibility.

Another capillary force often used in the formation of layers and other structures is the surface tension. Considerable control of the flow of suspensions can be obtained by controlling the surface tension.

These two 'tools' available to the suspension technologist will be discussed in more detail in the next two sections. These will be followed by some examples of the way in which capillary forces can be manipulated in production processes.

The stability of suspensions

The title photograph shows two bottles that contain equal amounts of finely ground quartz. In settling out of a suspension the quartz particles in the bottle on the left have formed a thick layer; on the right the layer is thinner and thus evidently more compact. The reason for this is that the particles on the right have settled to the bottom from a stable suspension. On the left the electrical double layer at the surface of the particles has been compressed by the addition of a small quantity of salt to such an extent that the stability of the suspension, which was due to double-layer repulsion, has been lost. This enabled a loose flocculated deposit to form. If a porous or compact layer of particles is required for a particular application, it is therefore possible to try to control the structure of the layer by manipulating the stability of the suspension from which the layer is to be made (by sedimentation, spraying, immersion, centrifuging or by other techniques). It is therefore no coincidence that a photograph of these same two bottles appeared in our journal more than a quarter of a century ago. They appeared in an article by S. A. Troelstra on coating by electrophoretic deposition [4]

[1] I. Langmuir, J. chem. Phys. 6, 873, 1938.
[2] E. W. Fischer, Kolloid-Z. 160, 120, 1958.
F. Th. Hesselink, A. Vrij and J. Th. G. Overbeek, J. phys. Chem. 75, 2094, 1971.
[3] B. Deryagin and L. Landau, Acta physicochim. 14, 633, 1941.
E. J. W. Verwey and J. Th. G. Overbeek, Theory of the stability of lyophobic colloids, Elsevier, Amsterdam 1948.
[4] S. A. Troelstra, Philips tech. Rev. 12, 293, 1950/51.

— one of the technologies that he helped to introduce at Philips for depositing small particles from a suspension. The bottles have not been faithfully preserved purely because of a feeling for tradition. They are brought out time and again to demonstrate the difference in sediment volume of layers formed from stable and unstable suspensions.

A stable suspension gives a compact deposit. The unstable suspension gives loose flakes that settle rapidly to form a thick porous layer. When it is shaken a porous layer of this type will break up easily and is then dispersed through the liquid in the form of small flakes. The thinner layer, however, that settled from the stable suspension is so compact that it is hardly affected by agitation of the liquid above it. The sediment from a stable suspension is therefore very difficult to disperse by shaking and stirring. In rheological terms, concentrated stable suspensions can be described as Newtonian liquids; with increasing concentration their viscosity can reach very high values. A concentrated *unstable* suspension often has the property of 'thixotropy'; quicksand is a typical example of a suspension that becomes more fluid the more vigorously it is shaken or stirred.

Why entirely different layers are precipitated from stable and unstable suspensions can be explained by the following reasoning. If a suspension is unstable — or if it becomes unstable in the process of drying up, for example — then every individual particle attaches itself to another one at the time and place where they come into contact. There is no system in the way in which particles come into contact, so that a haphazard network of cohering particles is formed. The voluminous sediment from an unstable suspension is a single continuous network of aggregated particles, a 'standing' floc. On shaking or stirring, forces are exerted on the whole of this large floc that are capable here and there of tearing the aggregates apart again at weak points. The continuous structure is destroyed, the smaller flocs become fairly mobile again relative to one another, and the liquid that was enclosed between them can flow freely again. All this together results in the thixotropic behaviour of the concentrated unstable suspension.

When a stable suspension is dried or goes into sedimentation the particles are systematically packed together. In relation to each other the particles remain mobile, so that they all remain moving until they have reached the position with the lowest attainable potential energy. Under the influence of gravity or of the surface tension of the drying suspension, such a sediment will therefore in general continue to settle until the particles have reached the most closely packed state. In this case, then, there is no 'standing' floc but a compact mass of individual particles. A layer of this type is not much affected by flow, which would have to penetrate between the closely packed particles to tear them apart.

If, in a particular production process, the particles of a suspension are to be located at a specific place by causing the suspension to flow, it is essential to take into account the property mentioned above. In pumping a stable suspension it is important that there should be no blind corners in the pipes, for this is where compact clumps can form.

Surface tension

For 'transport' purposes in thin layers the surface tension of a suspension can be used, e.g. for distributing particles evenly over a substrate. As a rule the surface tension of a suspension is equal to that of the medium in which the particles are dispersed. This surface tension determines which materials will be wetted by the suspension with a continuous thin layer, and from which materials the suspension, at a given layer thickness, will withdraw in the form of spherical droplets. A measure of the degree of wetting is the contact angle θ [5], which is found at the point of intersection of the interfaces between a solid substrate (S) and a liquid (L), between the substrate and a gas (G) or another liquid, and between the liquid and the gas (*fig. 2*). Along each of the three interfaces the surface tensions γ_{SL}, γ_{SG} and γ_{LG} exert an attractive force that tries to keep its own interface as small as possible. Mechanical equilibrium exists if

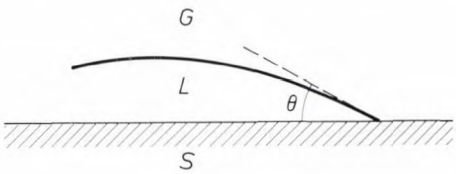
$$\gamma_{SG} = \gamma_{SL} + \gamma_{LG} \cos \theta.$$

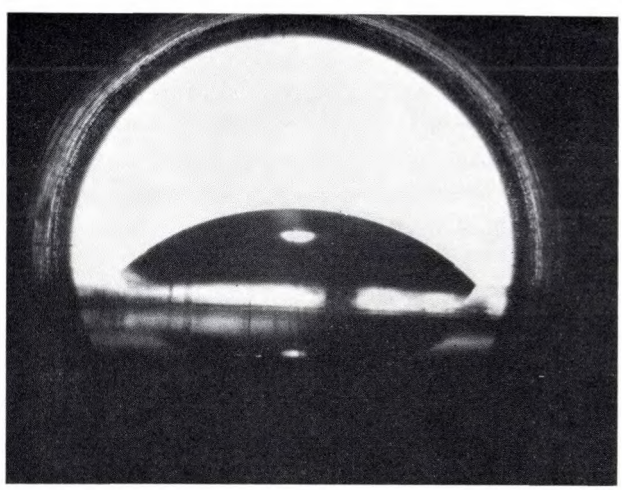
When $\gamma_{SG} \geqslant \gamma_{SL} + \gamma_{LG}$, there can be no equilibrium ($\theta = 0$). The boundary line where the three phases come together continues to move until the whole substrate has been covered by the liquid. In the case of a large contact angle, however, the liquid will roll over the substrate in the form of separate globules, and cohesion will often be very poor. Tissues or substrates with a low surface tension γ_{SG} are water-repellent, because γ_{LG} for water is relatively large. To wet such substrates well with an aqueous solution — to wash it, for example, or to let water penetrate into the pores — a surfactant (e.g. soap) is added, which lowers γ_{LG} and γ_{SL} by being adsorbed at the appropriate interfaces.

In pores, narrow tubes and capillaries a liquid that thoroughly wets the walls will rise. This effect of capillary rise is described by the equation $h = 2\gamma_{LG}/dga$ where it is assumed that $\theta = 0$. The quantity h is the height to which the liquid of density d can be raised by the surface tension γ_{LG} against the acceleration due to gravity g in a tube of radius a. In tubes with a diameter

of the order of 1 mm the capillary rise is about 1 cm — nothing very remarkable. But in a layer with particles of about 1 μm the cavities are of the same order of magnitude, and in such pores the surface tension is capable of raising fluids in thin films to a height of several metres. This would not be possible with a pump, because the layers are thin and highly viscous. In thin layers the surface tension has a much greater effect on the flow of the suspension than could be achieved with a pump or by making use of gravity.

In the production processes at Philips it is usually a question of depositing particles from a suspension in a layer of the desired thickness, strength, structure, porosity or packing. As we have seen, the suspension technologist can make use of the capillary forces for conveying the particles to their place and keeping them there.





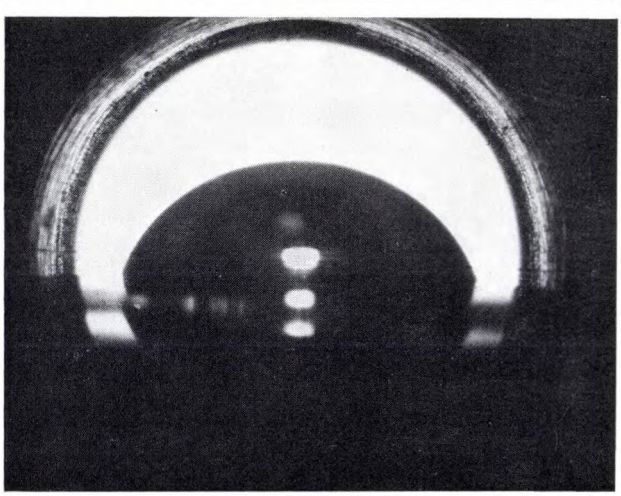


Fig. 2. The 'contact angle' θ, which occurs where three interfaces intersect (between a solid substrate S and a liquid L, between the substrate and a gas G, and between the liquid and the gas), is a measure of the extent to which the substrate can be wetted. For a given substrate material the contact angle depends on the various surface tensions of the liquids.

Electrophoretic deposition

Electrophoretic deposition, also known as cataphoresis, is used for applying emitter particles to cathodes, for applying latex globules to the jigs for rubber gloves and also, in a rather more complicated form, in electrophotography.

It seems a simple technique: electrically charged particles are deposited on a substrate that acts as an electrode with a polarity opposite to that of the particles. But there is more to it than that: for the particles to be deposited the suspension must be unstable in the neighbourhood of the electrode. The particles do in fact travel to the electrode in the electric field independently of one another, just as they do when settling from a suspension under the influence of gravity. If the suspension remained stable right up to the electrode, a compact layer would form there that was very viscous but with no internal cohesion. Such a layer would 'drip' from the electrode in tough droplets and collect in the vessel underneath. A process like this might well be useful for strongly concentrating a stable suspension (electro-decanting), but the electrode would not be coated with a layer of particles adhering to it.

To give good electrophoretic deposition the particles must coagulate at the electrode [6]. For this a suspension is used that is only just stable, and the process is arranged in such a way that electrolysis products formed by the passage of current through the electrode surface will cause flocculation at that surface.

Electrophoretic coating will thus give rise to a fairly porous deposit — typical of an unstable suspension — from a system that is apparently stable. In many cases, however, a porous deposit is just what is wanted. An example is to be found in the manufacture of large-area electrodes for batteries. *Fig. 3* shows an electrophoretically made electrode of $LaNi_4Cu$ powder for a rechargeable hydrogen storage battery [7]. In this case a second property of cataphoresis, which is of benefit here, has been put to use. At places where the current density is highest the most material is deposited. Thus, the largest amount of active electrode material will be available at the places where the currents will be highest while the battery is being charged and discharged.

[5] See for example C. A. Smolders, Rec. Trav. chim. Pays-Bas **80**, 650, 1961.

[6] H. Koelmans, Thesis, Utrecht 1955 (also published in Philips Res. Repts. **10**, 161, 1955).

[7] P. A. Boter, paper read at Novel Electrode Materials Symp., Brighton 1975 (not published).

Production of magnetic tape

The stability of a suspension cannot always be so easily manipulated as in the case just discussed. It is very difficult, for example, if not impossible, to stabilize a concentrated suspension of ferromagnetic particles. The magnetic attraction is so much stronger than the forces between 'neutral' particles that there is no known repulsive force that could compensate it. The main reason for this is that the range of magnetic fields in non-magnetic media is so much greater than that of other capillary effects. As a consequence the suspensions used in the production of magnetic tape (lacquers with about 10% by volume of magnetic pigment) are always flocculated when the suspension is at rest and thixotropic when it is agitated.

To form the magnetic tape it is necessary to deposit from such a suspension the thinnest possible layer of magnetic material with a high pigment volume concentration — 40 to 50% by volume. The needle-shaped pigment particles must be arranged in the layer like 'matches in a box', since this arrangement will give the strongest signal, the weakest noise, the highest sensitivity and the broadest frequency range. The tape also wears least when the volume fraction of pigment is high — but not too high.

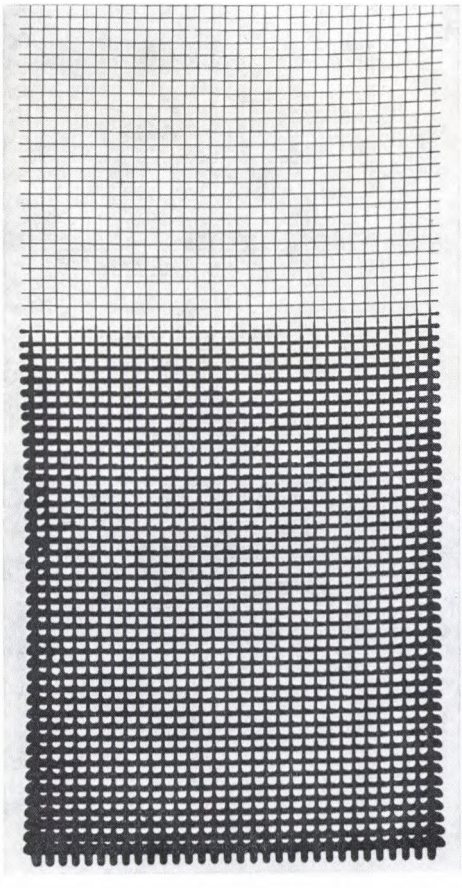


Fig. 3. An electrode of $LaNi_4Cu$ powder for a rechargeable battery, made by means of electrophoretic deposition. Most material has been deposited at the edges of the electrode where the current density was greatest. This has the advantage that where the currents are greatest in the operating battery there will also be the largest quantity of active electrode material. The upper part of the electrode was not immersed in the bath.

Much therefore depends on covering the magnetic particles with a layer in such a way that a stable suspension can be made from them. With such a suspension a closely packed pigment layer could be made instead of the standing porous floc of magnetic Fe_2O_3, CrO_2 or Fe that would otherwise be produced on pouring out a lacquer.

Dispersing agents are therefore added to the lacquer to envelope the magnetic particles in a coating of adsorbed material. They have a dual function: to ensure that the liquid lacquer can penetrate between the particles and remain there in pores (wetting) and to keep a spacing between the particles (stabilization). In this way the net attraction between the particles is sufficiently reduced for the separate particles to be kept apart from each other, at any rate on agitation, as in a stable suspension.

Although the adsorbed layer is unable to prevent the suspension from flocculating as soon as the flow stops, it does ensure that the lacquer will not rigidify while it is being poured, and that the particles in a standing floc formed in the poured layer can still readily be moved in relation to one another. When this has been achieved the magnetic forces — which had proved so troublesome — can even be turned to advantage, for it is now possible to orient the particles parallel with each other in an external magnetic field. Since the surface tension makes the layer with the oriented particles shrink, the particles are drawn towards each other like well wetted hairs. In this way a reasonably compact pigment layer is produced, by a circuitous route, from a highly unstable suspension.

Production of carbon potentiometers; critically filled layers

Another product in which the volume fraction of the 'pigment' in a poured layer determines the material properties is the carbon potentiometer. In a potentiometer with a total resistance of 1 MΩ the carbon layer poured from a suspension contains 6% of conducting graphite particles. For a resistance of 1 kΩ a layer with 30% of graphite is required. It is not only the electrical resistance of a potentiometer that is important, but also its resistance to wear and its ability to withstand the action of solvents and moisture. It is found that the best properties in these respects are obtained with a particular pigment volume concentration. This can be better understood if we take a closer look at the process of producing such a layer.

A carbon potentiometer is made by allowing a suspension of graphite or carbon black that contains a polymer as a binder to drip on to a substrate. After drying the suspension and thermal hardening of the binder a layer is obtained that is electrically conducting because of the contact between the carbon particles. The polymer medium in which the particles are embedded does not conduct.

The wear resistance of the potentiometer and its ability to resist corrosion are primarily properties of the polymer. With a small volume concentration of pigment in the polymer layer the particles have a mechanical strengthening function — like carbon black in car tyres. The wear resistance of the layer therefore initially increases with the volume of pigment.

To answer the question of why the wear resistance decreases above a 'critical pigment volume concentration' (CPVC), it is necessary to remember that most carbon suspensions from which potentiometers are made are unstable. Immediately after the pouring process a carbon layer forms with a high sediment volume, in which the interstices between the particles are filled with polymer and solvent. The solvent evaporates, leaving the polymer behind to preserve cohesion in the layer. The higher the pigment concentration in the suspension, the thicker the layer and the larger the volume of pores. The CPVC in the layer is that at which there is just enough polymer present to fill up the pores between the particles after the solvent has evaporated. With a higher PVC the material is thus slightly porous, and this degrades its resistance to wear and corrosion.

It follows from the above that the critical concentration of the pigment in the layer depends on the stability of the carbon suspension. Since a high CPVC is wanted, potentiometers are therefore best made from stable suspensions. To allow potentiometers of different resistance values to be made from layers of constant CPVC, the variation in the quantity of graphite is compensated by an equally large quantity of non-conducting particles acting as filler. An understanding of the connection between sediment volume and suspension stability has led to an appreciable improvement in the quality of this product.

Application of the coating in fluorescent lamps

To cover large areas evenly with particles, a stable suspension is caused to run down a wall, while the liquid evaporates from the suspension. This is the method used for applying the fluorescent coating in tubular fluorescent lamps. The suspension layer is applied to the tube wall by immersion, spraying or shaking, and it flows down the wall under its own weight. At the top of the tube no further suspension is added; suspension is however removed towards the bottom. Thus, the coating at the top end is always thinner than at the lower end, and it will become thinner the more suspension flows downwards. The art is to stop the thinning of the suspension layer once a particular thickness has been reached.

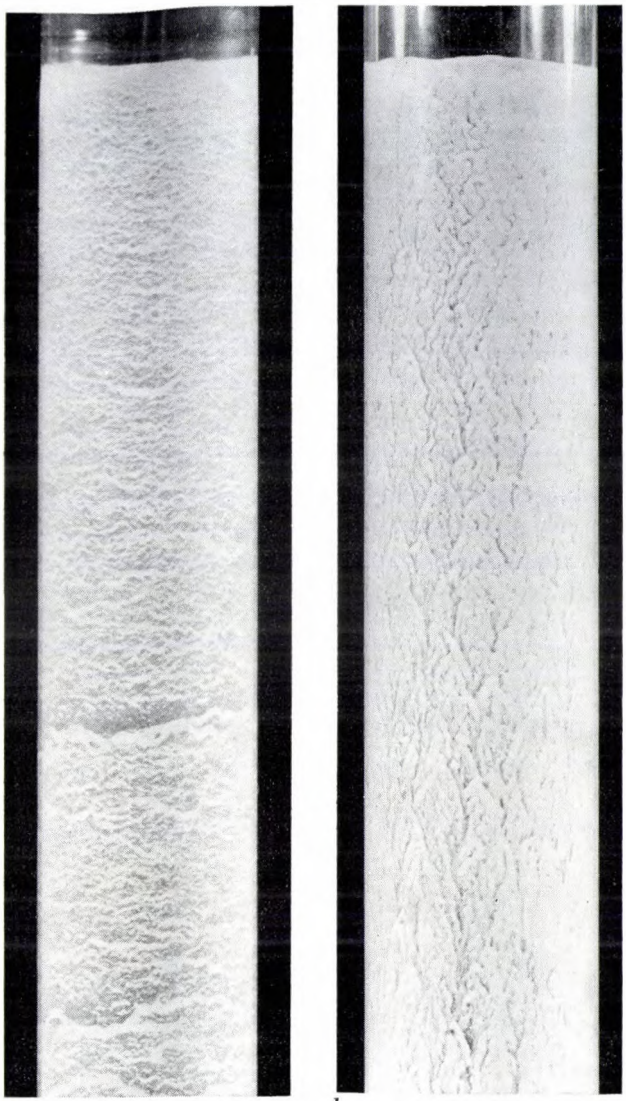
If the suspension layer somewhere near the top of the tube has become thinner than the diameter of a (wetted) particle in the suspension there, the particle will protrude from the liquid. This is associated with an enlargement of the liquid surface, which is counteracted by the surface tension of the suspension. The surface tension tends to minimize the curvature in the surface, thus ensuring that the particle remains as deep as possible in the liquid and is thus pressed firmly against the wall.

The particle continues to be surrounded by the liquid, with a curved meniscus towards the particle, like water against the walls of a wetted capillary. Smaller particles now arriving are held up in the stagnating volume of liquid and accumulate around the fixed larger particles. Finally all these particles are drawn firmly together by the liquid between them, and a more or less horizontal ring of particles forms. The process is repeated lower down and the wall takes on the appearance of a slightly clouded sky.

If the suspension also evaporates while flowing down the wall, a more pronounced clouding occurs; see *fig. 4a*. The stranded particles then behave like capillaries and draw up the liquid to compensate for the evaporation loss. We have observed that the liquid in an evaporating suspension becomes stationary as soon as the largest particles start to protrude above the surface. It is even possible for a vertical transport of liquid to take place from the bottom upwards. The capillary forces draw the suspension upwards (with the small particles in it) from places where the liquid layer is still thick to places where the layer is already so thin that particles have become firmly attached to the wall.

In this way the number of firmly adhering particles grows from the bottom upwards, and the adhering layer always forms with the same thickness, corresponding to the diameter of the largest particles. The tube of a fluorescent lamp is thus covered over the whole length with a layer of fluorescent particles of roughly constant thickness. When the volume fraction of particles in the suspension is large enough, the combination of downward flow, evaporation and capillary rise results in a dense and unbroken 'cloud' cover. The light from well made fluorescent lamps thus seems to come from 'heavy cloud'.

For this method of fabrication the surface tension, the viscosity, the particle size and the rate of evapora-



a *b*

Fig. 4. *a*) During the evaporation of the liquid horizontal 'clouds' of particles are deposited from a stable suspension on a vertical wall. *b*) The 'buttermilk effect' in the coating on the wall of a fluorescent lamp, which occurs when a non-stable suspension is used.

tion must be matched to one another to obtain the desired dense stacking. Here again, however, due attention must be paid to the stability of the suspension. Unstable suspensions give the 'buttermilk effect' (fig. 4*b*). The horizontal cloud formation due to capillary action does not then appear, but a flow region of meandering 'rivulets' between the standing flocs of the non-stable suspension. Owing to the thixotropic behaviour of the non-stable suspension the viscosity is low where the liquid flows, and the particles remain together in flocculated form where the liquid is stationary. A thick porous layer forms, traversed by deep flow gullies; pretty but undesirable. This illustrates the fact that suspension technology begins and ends with the ability to control the stability of suspensions.

Summary. In suspension technology the stability of suspensions and the surface tension are important aids in controlling the formation of layers. After an introductory discussion of these quantities, with special reference to the means of manipulation they offer, a number of examples are given of their use in Philips production processes. Descriptions are given of the fabrication of $LaNi_4Cu$ powder electrodes with a suspension that is only barely stable, the production of magnetic tapes from an intrinsically unstable lacquer suspension, made sufficiently stable for use by means of dispersing agents, the manufacture of carbon potentiometers with a high pigment volume concentration, and the manufacture of close-packed coatings on the wall of fluorescent lamps by appropriate control of surface tension, viscosity, particle size, evaporation rate and suspension stability.

Recent 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*
Mullard Research Laboratories, Redhill, Surrey, England	*M*
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*
MBLE Laboratoire de Recherches, 2 avenue Van Becelaere, 1170 Brussels (Boitsfort), Belgium	*B*
Philips Laboratories, 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A. (by contract with the North American Philips Corp.)	*N*

Reprints of most of these publications will be available in the near future. Requests for reprints should be addressed to the respective laboratories (see the code letter) or to Philips Research Laboratories, Eindhoven, The Netherlands.

G. A. Acket, W. Nijman, R. P. Tijburg & P. J. de Waard: Characterization of GaAs-AlGaAs double-heterojunction laser structures using optical excitation.
Gallium arsenide and related compounds, Proc. Conf. Deauville 1974 (Inst. Phys. Conf. Ser. No. 24), pp. 181-191; 1975. *E*

R. N. Bhargava: Recent advances in visible LED's.
IEEE Trans. **ED-22**, 691-701, 1975 (No. 9). *N*

A. Bril, J. L. Sommerdijk & A. W. de Jager: On the efficiency of Yb^{3+}-Er^{3+} activated up-conversion phosphors.
J. Electrochem. Soc. **122**, 660-663, 1975 (No. 5). *E*

L. K. Brundle: Experimental study of magnetodynamic edge-guided waves on a microwave ferrite substrate.
IEEE Trans. **MAG-11**, 1282-1284, 1975 (No. 5). *M*

T. M. Bruton: Investigation of the main structural units in Bi_2O_3-TiO_2 solutions.
J. Mat. Sci. **10**, 1798-1801, 1975 (No. 10). *M*

E. Constant*, A. Mircea, J. Pribetich* & A. Farrayre (* Université de Lille): Effect of transferred-electron velocity modulation in high-efficiency GaAs IMPATT diodes.
J. appl. Phys. **46**, 3934-3940, 1975 (No. 9). *L*

P. Delsarte & J. M. Goethals: Unrestricted codes with the Golay parameters are unique.
Discrete Math. **12**, 211-224, 1975 (No. 3). *B*

W. F. Druyvesteyn, A. W. M. van den Enden, F. A. Kuijpers, E. de Niet & A. G. H. Verhulst: Magnetic bubbles.
Solid state devices, 1974, Proc. Conf. Nottingham (Inst. Phys. Conf. Ser. No. 25), pp. 37-74; 1975. *E*

W. G. Essers, G. W. Tichelaar, G. J. P. M. van den Heuvel & G. Jelmorini: Welding in thermally ionised gas.
Advances in Welding Processes 3rd Int. Conf., Harrogate 1974, Vol. I, pp. 113-120. *E*

A. Farrayre, B. Kramer & A. Mircea: High efficiency avalanche diodes from 10 to 15 GHz.
Gallium arsenide and related compounds, Proc. Conf. Deauville 1974 (Inst. Phys. Conf. Ser. No. 24), pp. 71-76; 1975. *L*

Y. Genin & Y. Kamp: Counterexample in the least-squares inverse stabilisation of 2D recursive filters.
Electronics Letters **11**, 330-331, 1975 (No. 15). *B*

G. J. van Gurp: Resistivity, grain size, and structure of vacuum-deposited Co films.
J. appl. Phys. **46**, 1922-1927, 1975 (No. 5). *E*

C. M. Hargreaves: Electromagnetic proximity effects and their consequences for radiation shielding.
Proc. 5th Int. Cryogenic Engng. Conf., Kyoto 1974, pp. 574-577. *E*

E. E. Havinga: Influence of repulsive energy on structural parameters of close-packed metal structures.
J. less-common Met. **41**, 241-254, 1975 (No. 2). *E*

J. C. M. Henning & J. H. den Boef: Strain-modulated electron spin resonance spectra of $3d^5$ ions in ZnSe.
Phys. Stat. sol. (b) **69**, 153-161, 1975 (No. 1). *E*

W. J. van den Hoek (Philips Lighting Division, Eindhoven): An evaluation of the carbon-filament incandescent lamp.
Philips Res. Repts. **31**, 129-152, 1976 (No. 2).

L. Hollan: Vapour phase epitaxial growth of GaAs for multiple applications.
Gallium arsenide and related compounds, Proc. Conf. Deauville 1974 (Inst. Phys. Conf. Ser. No. 24), pp. 22-30; 1975. *L*

A. P. Hulst & J. O. Zwolsman (Philips Elcoma Division, Eindhoven): Adaptive control in resistance welding small components.
Advances in Welding Processes 3rd Int. Conf., Harrogate 1974, Vol. I, pp. 27-36. *E*

H. D. Jonker: Investigation of the phase diagram of the system PbO-B_2O_3-Fe_2O_3-Y_2O_3 for the growth of single crystals of $Y_3Fe_5O_{12}$.
J. Crystal Growth **28**, 231-239, 1975 (No. 2). *E*

D. Kasperkovitz: A 20-MHz p-n-p-n shift register with current mirror coupling.
IEEE J. **SC-10**, 125-129, 1975 (No. 3). *E*

E. Kauer & T. Thalhammer: The potential of solar energy.
Atomkernenergie **25**, 162-172, 1975 (No. 3). *A, E*

M. Klinck: Simulation aided modeling of the dynamic behaviour for some elements of a surface water treatment plant.
Modeling and simulation of water resources systems, editor G. C. Vansteenkiste, North-Holland Publ. Co., Amsterdam 1975, pp. 467-475. *H*

J.-P. Krumme & H. J. Schmitt: Ferrimagnetic garnet films for magnetooptic information storage.
IEEE Trans. **MAG-11**, 1097-1102, 1975 (No. 5). *H*

D. E. Lacklison, G. B. Scott, R. F. Pearson & J. L. Page: Magneto-optic displays.
IEEE Trans. **MAG-11**, 1118-1120, 1975 (No. 5). *M*

J. Matthys: Linear vector recurrences.
Philips Res. Repts. **31**, 120-128, 1976 (No. 2). *B*

A. R. Miedema, R. Boom* & F. R. de Boer* (* Univer siteit van Amsterdam): On the heat of formation of solid alloys.
J. less-common Met. **41**, 283-298, 1975 (No. 2). *E*

A. Mircea-Roussel & F. Rondelez: Pretransitional behavior of conductivity and dielectric properties above a smectic *C* - nematic transition in 4-4′-di-(*n*-alkyloxy)azoxybenzenes.
J. chem. Phys. **63**, 2311-2316, 1975 (No. 6). *L*

B. H. Newton, R. E. Pearson & J. C. Williams: The design and realisation of M.I.C. subassemblies on ferrite substrates.
IEEE Trans. **MAG-11**, 1270-1272, 1975 (No. 5). *M*

T. G. J. van Oirschot, G. A. Acket & W. J. Bartels: LPE growth and properties of GaP-(AlGa)As-GaAs heterostructures as a function of the GaP substrate roughness.
J. appl. Phys. **46**, 1893-1900, 1975 (No. 5). *E*

R. C. Peters & A. T. Vink: On the oxygen donor concentration in GaP grown from a solution of GaP in Ga.
Gallium arsenide and related compounds, Proc. Conf. Deauville 1974 (Inst. Phys. Conf. Ser. No. 24), pp. 254-265; 1975. *E*

C. Piaget, P. Guittard, J. P. André & P. Saget: Photoemission and secondary emission of gallium phosphide epitaxial layers.
Gallium arsenide and related compounds, Proc. Conf. Deauville 1974 (Inst. Phys. Conf. Ser. No. 24), pp. 266-274; 1975. *L*

H. Rau: Thermodynamics of dense phosphorus vapours.
J. chem. Thermodyn. 7, 903-912, 1975 (No. 10). *A*

J. M. Robertson, C. J. M. Rooymans & W. Kamminga: YIG single crystals flux-grown under oxygen and argon pressure.
High Temp. - high Press. **6**, 609-613, 1974 (No. 6). *E*

P. Röschmann: Separation of anisotropy and porosity contributions to inhomogeneous broadened FMR linewidth in polycrystalline YIG.
IEEE Trans. **MAG-11**, 1247-1249, 1975 (No. 5). *H*

G. B. Scott, D. E. Lacklison, H. I. Ralph & J. L. Page: Magnetic circular dichroism and Faraday rotation spectra of $Y_3Fe_5O_{12}$.
Phys. Rev. B **12**, 2562-2571, 1975 (No. 7). *M*

J. G. Siekman, Th. M. B. Schoenmakers, K. N. Eindhoven* & D. van der Torre* (* Metal Institute TNO, Apeldoorn): New microprobe techniques for measuring the current distribution in an electron beam used for welding.
J. Physics E **8**, 391-395, 1975 (No. 5). *E*

J. L. Sommerdijk & A. Bril: Green luminescence of $CsCaF_3:Eu^{2+}$.
J. Luminescence **10**, 145-147, 1975 (No. 2). *E*

W. Tolksdorf: Growth and properties of garnet films for storage application.
IEEE Trans. **MAG-11**, 1074-1078, 1975 (No. 5). *H*

H. Ton: Physical properties of the plasma-MIG welding arc.
J. Physics D 8, 922-933, 1975 (No. 8). *E*

J. D. B. Veldkamp & R. J. Klein Wassink (Philips Centre for Technology, Eindhoven): Grindability of brittle materials: a theoretical and experimental investigation.
Philips Res. Repts. **31**, 153-189, 1976 (No. 2). *E*

C. H. F. Velzel: Inverse Fourier Spectroscopy, Part II: Experiments.
Philips Res. Repts. **31**, 97-119, 1976 (No. 2). *E*

J. F. Verwey, R. P. Kramer & B. J. de Maagt: Mean free path of hot electrons at the surface of boron-doped silicon.
J. appl. Phys. **46**, 2612-2619, 1975 (No. 6). *E*

P. J de Waard: Calculation of the far-field halfpower width and mirror reflection coefficients of double-heterostructure lasers.
Electronics Letters **11**, 11-12, 1975 (No. 1). *E*

J. M. Woodcock & D. J. Clark: The ion implantation of donors for n^+-p junctions in GaAs.
Gallium arsenide and related compounds, Proc. Conf. Deauville 1974 (Inst. Phys. Conf. Ser. No. 24), pp. 331-341; 1975. *M*

D. J. Zwanenburg & Th. A. M. M. Maas: The reactions of salicylaldehydes and 2,2-diaryl-2*H*-chromenes with 1,1-bis[*p*-(dimethylamino)phenyl]ethylene; photochromic chromenes.
Recueil Trav. chim. Pays-Bas **94**, 8-12, 1975 (No. 1). *E*

PHILIPS TECHNICAL REVIEW

VOLUME 36, 1976, No. 10

Control of the surface-water purification plant for the Amsterdam Water-Supply Authority

M. Klinck

The supply of drinking water, especially in urban areas, is a problem that goes back to the remote past, as witnessed by the many remains of water-piping systems left behind by the ancient Greeks and Romans. The city of Amsterdam, which has virtually no potable well water or ground water in its immediate vicinity, had originally to make do with water from its canals. When growing silt accretion and pollution made this no longer possible, the city had to rely on a special shipping fleet to bring in drinking water from nearby rivers. The water was sold by the bucket. In 1854, partly on the initiative of the author Jacob van Lennep, a dune-water pipeline was laid. When in 1889 the reserves of ground water in the dunes threatened to become insufficient to meet the city's needs, a second water-pipeline system was laid to distribute the river water previously brought in by ship. Until 1932 the two water pipelines existed side by side: the river water for scrubbing the steps, the dune water for human consumption. Subsequently the river-water pipeline was replaced by a pipeline drawing water from the Bethune polder, through the lakes in the Loosdrecht area. Both systems now supplied potable water, and could be combined to form a single system. Since 1957 the ground-water reserves in the dunes have been supplemented by infiltration with river water, raising the capacity of the dune-water supply to 70 million cubic metres a year, and a new surface-water treatment plant has been built, with a provisional capacity of 30 million cubic metres a year. This capacity can be supplemented by another 30 million cubic metres a year.

Reliable drinking water is one of man's most vital needs. In many places, especially in urban areas, this need is met by water-supply authorities, which distribute purified water through a mains system. Since mains water is not only used for direct consumption but also for many and various cleaning purposes and in industry, the water demand per head of population continues to increase. This rise in per capita requirements and the expected growth of population lead some authorities to predict an increase in demand of 30 to 50% in the next ten years [1]. To meet such an increased demand and to cope with increasing surface-water pollution, water-treatment and purification plants will have to be expanded and new plants built.

The raw water for the treatment plant is in most cases extracted from the ground water through a system of wells. Surface water from lakes, rivers or channels is only used where either the quality or the quantity of the ground water is insufficient. Ground water is filtered through sand beds, which usually ensures adequate purification.

The capacity of the ground-water wells is not sufficient to meet the growing demand, and an increase in the number of wells is ruled out by practical considerations. In this situation increasing use is being made of the surface water, but this is becoming more and more polluted by urban and industrial effluents. As a result increasingly complex treatment plants have to be used to purify this surface water.

In addition to their complexity, the large capacity of the surface-water treatment plants of some major cities makes it difficult to control the installations manually.

Dipl.-Ing. M. Klinck is with Philips GmbH Forschungslaboratorium Hamburg, Hamburg, West Germany.

[1] Het waterbeleid in België, H_2O 2, 185, 1969.

The number of stages to be controlled is so great that it has become very difficult to monitor the system as a whole, particularly since some of the processes are very slow. In a case such as this the obvious solution is automatic control of the complete installation.

We have designed a control system for a surface-water treatment plant belonging to the Amsterdam Water-Supply Authority [2]. The plant consists of two units, situated twelve kilometres apart. The measurement and control instruments are regularly monitored by one person in a central control room and there is no need for more than two supervisory staff. A small computer deals with the routine control, reports on the state of all sections of the plant and gives warning signals if any undesirable situation threatens to arise anywhere within the system.

Before dealing with the design of the actual control system we shall first look at the functions required of the system and at some of the factors that played a part in the choice of the design procedure. A description of the result then follows: a fully automatic control system, and its performance. The diagram below shows the principal parts of the treatment plant.

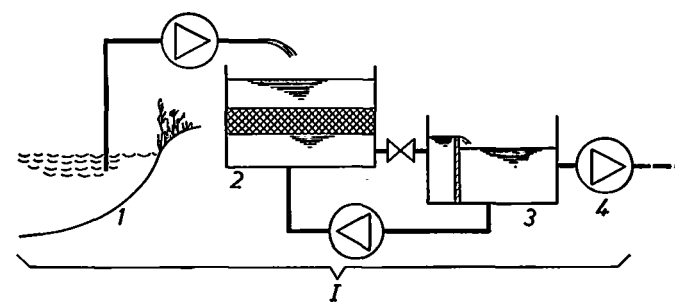
Functions required of the control system

The primary functions of the water-treatment control system are to ensure that there is a sufficient supply of water available at particular points (to cope with sudden peaks in demand), that basins do not overflow and that the pumps do not run dry or become overloaded. The system must also ensure that the filters operate correctly, keep the flow rates through the filters as constant as possible, and ensure that any

[2] M. Klinck, AICA Symp. on Hybrid computation in dynamic systems design, Rome 1974, section D1-1, p. 1.
M. Klinck, in: G. C. Vansteenkiste (ed.), Modeling and simulation of water resources systems, North-Holland, Amsterdam 1975, p. 467.

I Loenerveen Station

1. Storage lake. The main storage reservoir is a dammed-up part of the Loenerveense Plas, one of the lakes in the region between Amsterdam and Utrecht. This lake is fed from the reservoir water of the Bethune polder. The water is pumped by constant-speed centrifugal pumps, whose blades can be adjusted to regulate the flow. A number of pumps can work in parallel, depending on the demand.



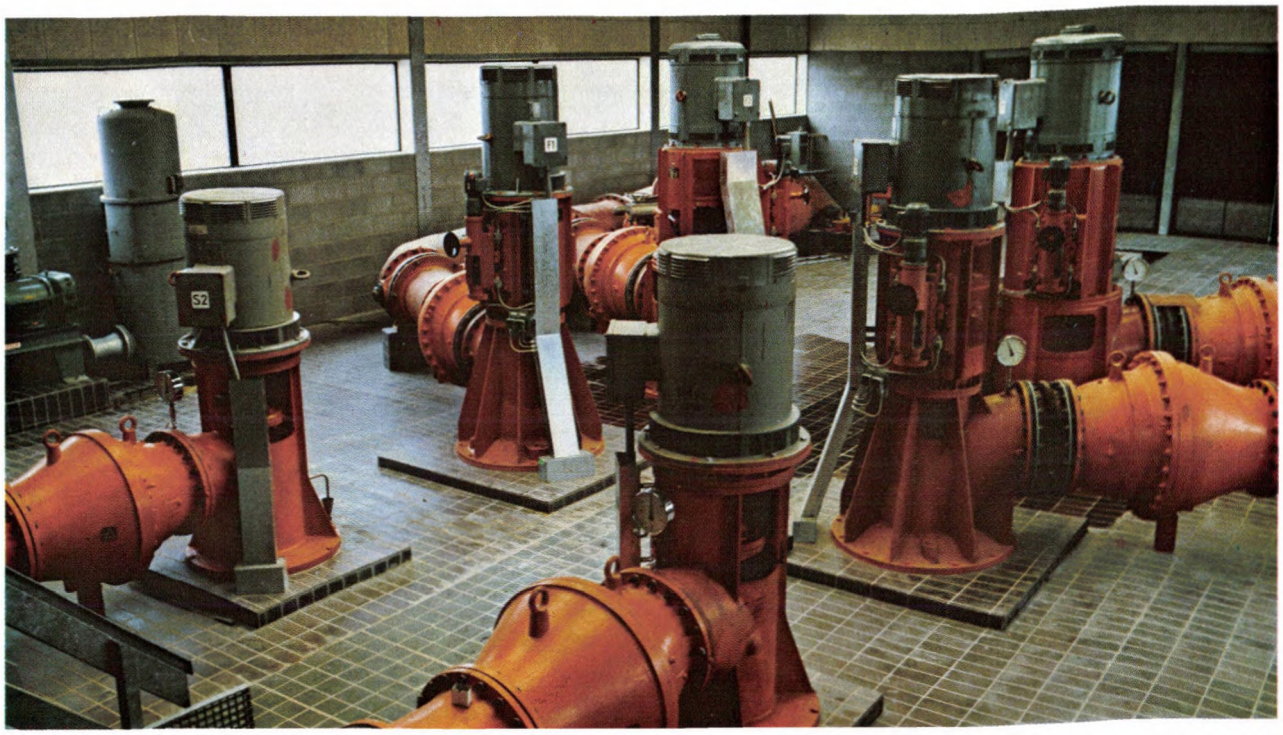
2. Rapid sand-filter section. There are 24 parallel filter basins, which, in groups of six, have common input and output pipes. Each filter basin is about 3 metres deep, with a sand bed about 1 metre thick about half way down. About every 24 hours each filter has to be cleaned by backwashing for 5 minutes with about 250 m^3 of water. The total outflow from the filters may vary from 1800 to 5300 m^3/h. The permissible static deviation from a nominal value for the water level in the filter basins is 5 cm, the dynamic deviation 10 cm.

3. Storage reservoirs. From these the filtrate water is pumped to the final treatment station at Weesperkarspel. These reservoirs also supply the water needed for the periodic flushing of the filters. The permissible static deviation of the water level is 10 cm, the dynamic deviation 15 cm.

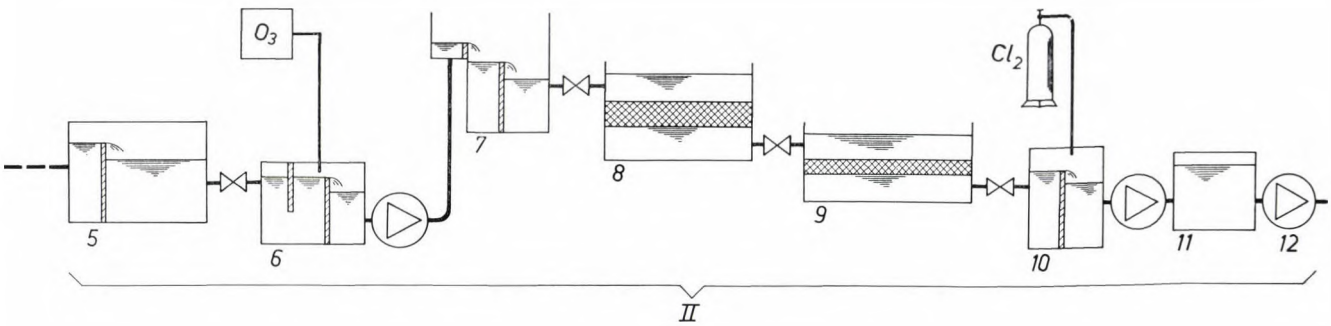
4. Pumping station and transport pipeline. The transport pipeline is 12 kilometres long and consists of two pipes. There are three constant-speed centrifugal pumps with adjustable blades and of different capacities. The pump switched in is the one capable of delivering the required flow. There is also a standby pump, with a capacity equal to the largest of the series of three.

II Weesperkarspel Station

5. Storage reservoir. The storage reservoir at the input of the final treatment station serves as a standby supply of drinking water in the event of a stoppage due to a breakdown in the main pipeline. The water level in the reservoir is 130 cm; deviations up to 15 cm are permitted.



Pumps at the Loenerveen station. The two pumps in the foreground are for filter backwashing, the four at the back for the transport pipeline.



6. OZONIZER STATION. In the ozonizer station the water is brought into contact with ozone produced on the spot by means of a high-voltage discharge. Impurities affecting the colour, smell or taste are oxidized and made harmless.

7. COAGULATION STATION. Coagulants are added here to bind solid impurities. An activated-carbon treatment is also given to improve the taste of the water. Coagulated impurities are removed by sedimentation.

8. RAPID SAND FILTERS. There are 12 of these filters; the design is much the same as that of the sand filters in the raw-water treatment at Loenerveen.

9. SLOW SAND FILTERS. Here the last solid impurities are removed from the water. Biological cleansing is also carried out by the bacterial flora on the sand layer. The flow velocity in this filter is low (0.5 m/h) and it must not change by more than 0.1 m/h per hour. The filters are cleaned every few months by removing 2 cm from the top layer of the sand. After this it takes some time for the bacterial flora to re-establish itself.

10. CHLORINATION STATION. Some chlorine is added to the purified water to counteract algae growth in the pipes of the mains network. Ideally, all the chlorine should just have been used up when the water runs out of the kitchen tap.

11. DRINKING-WATER RESERVOIRS. The purified water is stored in two reservoirs 52 m in diameter and 7 m high.

12. DRINKING-WATER PUMPS. These pumps convey the water from the reservoirs into the Amsterdam water mains. These pumps are not automatically controlled.

changes required in the filters take place at a sufficiently slow rate. Flow rates in the ozonization and chlorination basins must also remain as constant as possible to ensure optimum dosage by diffusion. The control system also has to ensure that where a number of pumps work in parallel the total flow is evenly distributed among these pumps.

The primary reference value for the complete control system is the quantity of drinking water to be delivered to the mains network. The reference values for the control of all parts of the process have to be derived from this primary value.

Design of the control system; simulation

Ideally the control system should be designed at the same time as the entire treatment plant. In this way the choice of valves, pumps, basin dimensions and pipe diameters can also take into account the requirements that these parts have to meet for the control system. On the other hand, in the design of the control system it is desirable to have some knowledge of the behaviour of the purification plant. There are various ways in which such knowledge can be acquired in the design phase. Use can be made of experience gained in a previous design of a similar plant, and comparable plants elsewhere can also be studied. If this cannot be done, however, some kind of model of the plant will have to be used to simulate its behaviour [3].

Simulation-aided design

To simulate a plant, whether or not it includes a control system, a scale model (mock-up) or a mathematical model can be used. Making a scale model can be difficult and expensive, but in some cases there is no alternative. In most cases, however, a mathematical model is the better choice. The model consists of mathematical equations in which the variables correspond to the variables in the plant to be simulated. Here again, two cases may be distinguished. The equations give the relation between input and output quantities of parts of the installation as found by measurement without expressing the physical laws underlying the behaviour ('black-box model'). In the other case the equations can be based on the physical laws themselves. This type of model will often be complicated, but it does give a better understanding of the backgrounds of the simulated behaviour.

Simulation enables the designer to compare various possible solutions of the same problem and thus, in a number of steps, to approximate the desired behaviour with sufficient accuracy. Once a plant, including the control system, has been designed in this manner, simulation then allows the behaviour of the plant to be studied under a variety of conditions. The knowledge thus gained proves very valuable when setting up the final version of the design and putting it into service.

Use of a mathematical model

Simulation on the basis of a mathematical model can be carried out in two ways. An analog computer can be used, i.e. a combination of operational amplifiers and passive networks whose behaviour is defined by the same equations that form the model. The currents and voltages in the analog correspond to the quantities in the plant to be simulated. An advantage of this analog computer is that it works rapidly: all the information is processed in parallel and all data are available at the same time. A disadvantage is that the currents and voltages in the analog may reach practical limits earlier than the corresponding quantities in the simulated plant. All the modifications in the model indicate modifications in the wiring of the analog, and this can be rather time-consuming.

The other possibility is to perform the simulation with the aid of a digital computer. Numerical values for the quantities of the plant to be simulated are then calculated from the equations in the model. There are now very few limitations to the magnitude of the numerical values. Changes in the model entail changes in the computer program. The information is now processed consecutively by the computer, which limits the speed. For this method of simulation special languages have been developed, such as COSILA (continuous system simulation language) and CSMP (continuous system modelling program).

It is also possible to combine the two simulation methods. This is done by interconnecting a digital computer and an analog computer by means of digital-to-analog converters to form a hybrid computer. Part of the plant is simulated by an analog, another part is simulated numerically. The division between the two simulation methods can be arranged to give the maximum benefit from the advantages of the two methods, yet with the minimum inconvenience from their disadvantages.

Behaviour of a rapid sand filter

As an example of the procedure we have adopted in simulating the water-purification plant the simulation of one of its components will be discussed: a rapid sand filter [4].

The filter consists of a concrete tank with a bed of sand about a metre deep at about half the height (*fig. 1*). A difference in level before and after the filter keeps the water flowing through the sand bed. The contamination of the sand bed will gradually increase

in use, and this causes the flow resistance to rise. Changing the setting of a valve in the output pipe will keep the flow constant in spite of the increasing contamination. When the contamination has risen too far, the filter is cleaned by backwashing for about five minutes, either with water or with an air/water mixture. Backwashing will usually be necessary once every 24 hours.

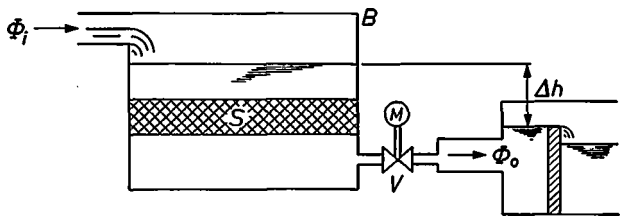


Fig. 1. Schematic cross-section of a rapid sand filter. The concrete filter basin B contains a sand bed S about 1 metre thick. The difference in level Δh keeps the water flowing through the sand bed. The butterfly valve V, driven by a motor M, keeps this difference in level constant, in spite of variations in the inflow Φ_i or the transmission of the sand bed. Φ_o outflow.

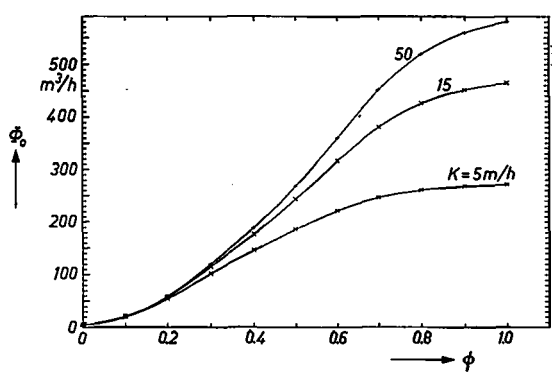


Fig. 2. Relation between the outflow Φ_o and the normalized valve setting ϕ for various values of the transmission K of the sand bed. The normalized valve setting is defined as $\phi = \alpha/\alpha_{max}$, where α is the actual valve setting and α_{max} is the position for maximum cross-sectional area of the opening.

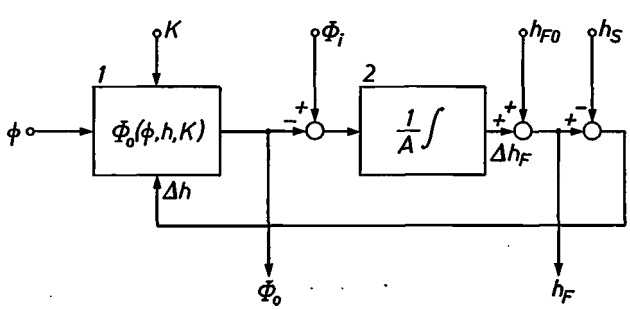


Fig. 3. Block diagram illustrating the dynamic behaviour of a rapid sand filter. Block *1* gives the outflow Φ_o as a function of the valve setting ϕ, the transmission K of the sand bed and the water-level difference Δh. Block *2* gives the change Δh_F in the water level above the filter, as found from the integral of the difference between Φ_o and the inflow Φ_i after dividing by the filter surface area A. Adding h_{FO}, the value of the initial water level, gives the water level h_F above the filter. The difference between h_F and the water level h_S in the storage basin gives the value of the water-level difference Δh to be fed back to block *1*.

Static behaviour

The equations that describe the static behaviour of the filter are relatively easy to formulate. The flow Φ_o through an opening of area A in the bottom of a tank with a water level h is given by:

$$\Phi_o = A\sqrt{(2gh)},$$

where g is the acceleration due to gravity. For an opening in the wall of the tank, h must be replaced by h_{eff}, the effective water level above the opening. All obstacles that the water has to pass on leaving the tank can be taken into account by subtracting various corrections from h_{eff}. In many cases these corrections will be a function of the flow Φ_o. For the valve in the output pipe this reduction in height is also a function of the valve setting. It is difficult to derive this function theoretically, and we therefore use a relation that approximates as closely as possible to the experimentally determined behaviour.

Once all the corrections to the water level have been taken into account, we can find a relation between the flow Φ_o and the valve setting ϕ. This relation is presented in *fig. 2* for various values of the transmission K of the sand bed. It is assumed in the calculation that the flow rate is controlled in such a way that the water level above the sand bed is constant.

Dynamic behaviour

To describe the dynamic behaviour of the filter we must take into account the effect of variations in the water level. These variations Δh are found by integrating with respect to time the difference between the inflow Φ_i and the outflow Φ_o:

$$\Delta h(t) = \int_0^t (\Phi_i - \Phi_o)\,d\tau.$$

The block diagram in *fig. 3* shows the filter as a dynamic system. The diagram consists of a block that represents the static behaviour with a feedback loop that contains an integration and a number of algebraic operations. This means that the filter can be described as a first-order system in which the transfer function and the time constant are functions of the valve setting ϕ.

To study the dynamic behaviour we used a hybrid computer, which allows a rapid display to be obtained of the effect of changes in the values of various param-

[3] Proc. UKAP 5th Control Conf., Modeling and simulation for applied control systems, Institute of Measurement and Control, University of Bath, England, 1973.
[4] D. W. Hilder, Proc. Conf. on Computer uses in water systems, Reading, England, 1973.
M. Klinck, Proc. IFAC Symp. on Control of water resource systems, Haifa 1973.
M. Klinck, Proc. Conf. on Computer uses in water systems, Reading, England, 1973, p. 101.

eters. *Fig. 4* shows the reaction of the filter to a sudden change in the valve setting. The final value both of the water level and of the flow depends on the transmission K of the sand bed. The time constant that gives the speed at which this final value is reached, however, is practically independent of K. All quantities are found to depend on the initial valve setting.

To obtain a better understanding of the factors that determine the behaviour of the filter it is useful to study a simplified model in which it is assumed that Φ_0 depends linearly on the valve setting ϕ, on the water level h and on the transmission K. Strictly speaking the results obtained with this model apply only to a single setting of the filter, but the results are still reasonably good in the immediate neighbourhood of this setting. The 'linearized model' is represented by the block diagram in *fig. 5*. The various quantities in this model are shown in *fig. 6* as a function of the valve setting ϕ. Values for $\phi < 0.2$ have not been calculated since they will hardly ever be encountered in practice.

Automatic control of the filter

The principal task of the control system for the rapid sand filter is to keep the water level above the filter constant in spite of variations in the inflow Φ_i. The actuating variable for this control is the setting of the valve in the output pipe. The valve setting can be altered by means of a constant-speed d.c. motor, operated with a three-position stepping controller (forward, stop, reverse). This controller reacts to the difference between the measured water level and the desired level.

This combination of a three-position stepping controller and a motor can be regarded as an analog PI

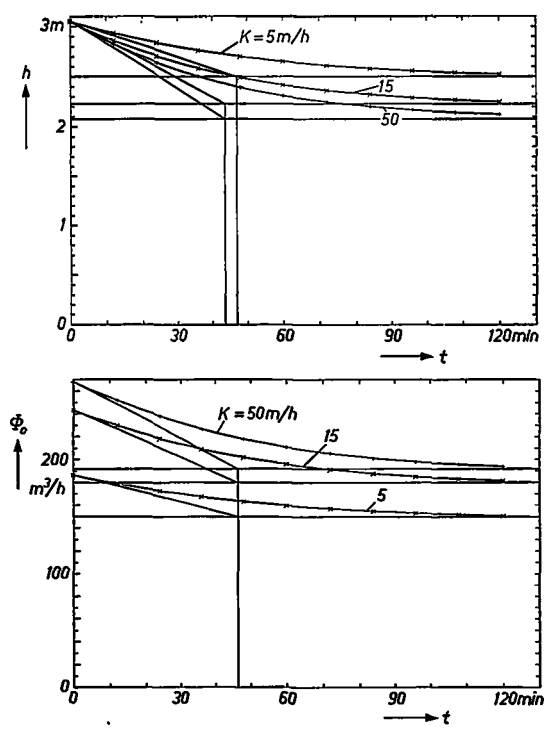


Fig. 4. Reactions of the hybrid computer model of a rapid sand filter to an abrupt change in the valve setting from $\phi = 0.4$ to $\phi = 0.5$. *Above:* Change in the water level h above the filter as a function of time t. *Below:* Change in the outflow Φ_0 as a function of t. Three values are always taken for the transmission K of the sand bed. When the filter is considered as a first-order system, the point of intersection of the tangent to a curve at $t = 0$ with the asymptote of the curve gives the time constant of the system. Almost irrespective of the value of K, this constant is always found to be about 45 minutes.

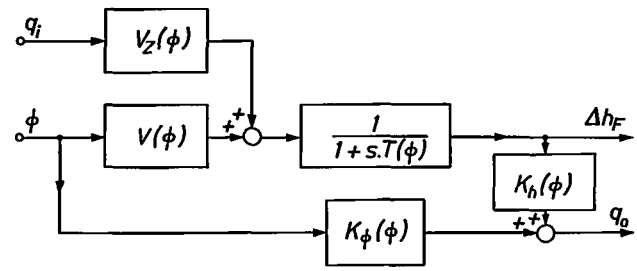


Fig. 5. Block diagram of the linearized model of a rapid sand filter. The changes Δh_F in water level and in the outflow q_o are functions of the inflow q_i and of the valve setting ϕ. The parameters V, V_z, K_h and K_ϕ, and also the time constant T are all affected by the valve setting. The function with the differential operator s provides a first-order delay for the block in which it appears.

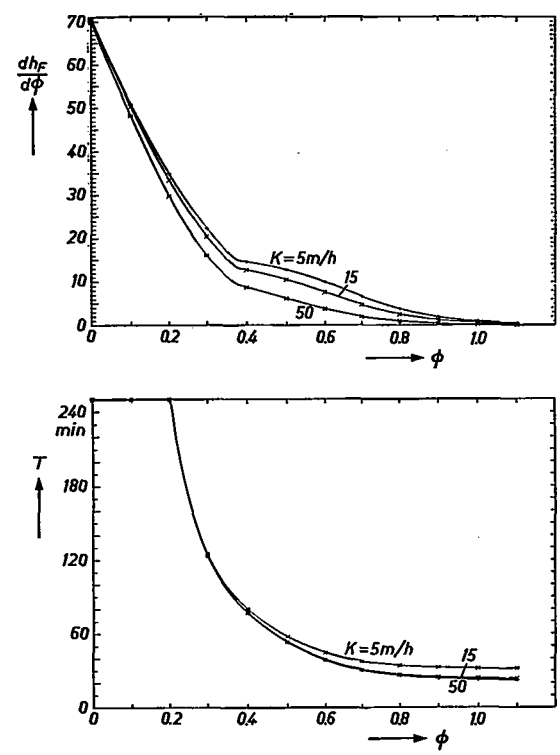


Fig. 6. Variation of two quantities for the linear model of a rapid sand filter as a function of the normalized valve setting ϕ. *Above:* The gain $dh_F/d\phi$. The bends in the curves arise because of the way in which the valve resistance is approximated as a function of ϕ. *Below:* The time constant T of the linearized system. In both cases the curve is calculated for three values of the transmission K of the sand bed.

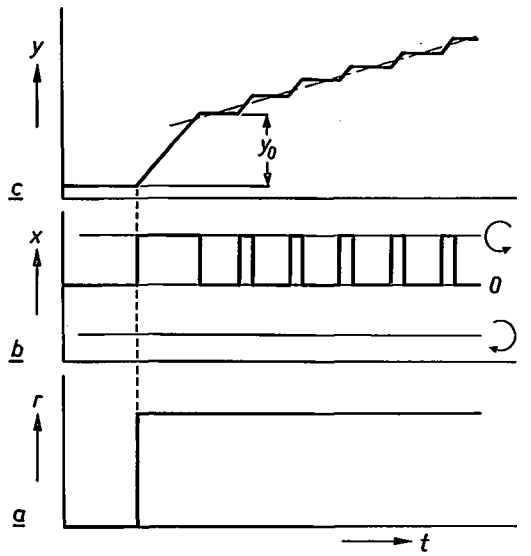


Fig. 7. Step response of a motor operated by a three-position stepping controller with feedback. *a*) The control signal r, with step response, as a function of time t. *b*) The excitation x of the motor, alternately stationary and rotating in one direction. *c*) The setting y of the valve operated by the motor. The step response consists of an initial displacement y_0 proportional to the abrupt change in r — proportional operation — and a change occurring on average linearly with time, where the slope is also determined by the magnitude of the change in r — integral operation.

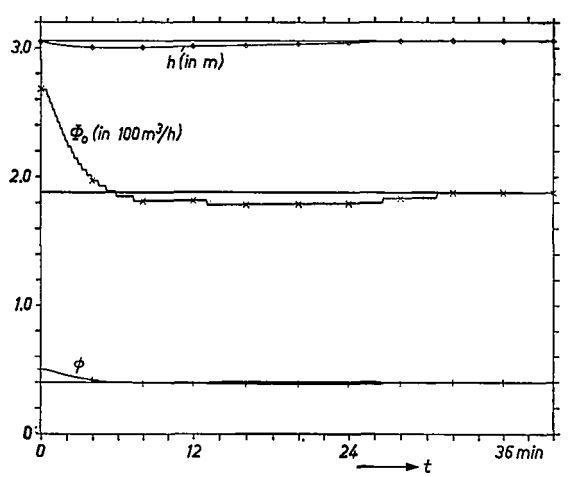


Fig. 8. Behaviour, as a function of time t, of a number of quantities of the real, nonlinearized model of a rapid sand filter at an abrupt decrease in the inflow. The quantity ϕ is the normalized valve setting in the outflow pipe. Φ_0 is the outflow. h is the water level above the filter.

— the deviations between the desired and actual water levels must not exceed a specified value;
— the rate of change of the flow rate of the water must not exceed a specified value;
— the static and dynamic behaviour of the filter should be independent of the outflow Φ_0.

The first two requirements are satisfied by an appropriate choice of the parameters K_p, τ and K_n. The behaviour of the filter with optimum values for these parameters is illustrated in *fig. 8*. It is not possible, however, to meet the last requirement with fixed values of the three parameters. The only solution would appear to be to keep adjusting the parameters to the valve setting ϕ by using adaptive control. Investigations have shown, however, that in the normal operating range the behaviour in practice is in fact acceptable without such refinements.

Control of the raw-water treatment

Elements of the system

The water from the storage reservoir is pumped to the rapid sand filters by centrifugal pumps, which operate at a constant speed. The position of the pump blades can be adjusted by a motor to regulate the flow.

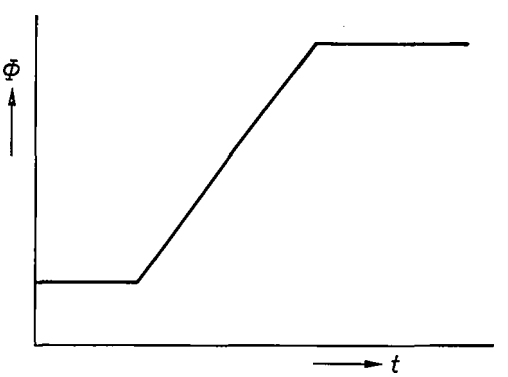


Fig. 9. Behaviour as a function of time t of the outflow Φ of a pump with adjustable blades at an abrupt change in the control signal. Because of the constant speed of the motor that adjusts the position of the blades, the output changes at a constant rate until the new value is reached.

controller, i.e. one that acts both as a proportional and as an integral controller (*fig. 7*).

Most parameters of the system will vary during operation, since they are a function of the valve setting ϕ. The only constants are the proportional transfer ratio K_p, the integration constant τ of the PI controller and the transfer ratio K_n of the water-level transducer. These constants can now be used for ensuring that the whole process meets the following requirements:

The flow of a pump is a nearly linear function of the position of the blades. The reaction of the flow of a single pump to a change in the reference value is shown in *fig. 9*.

When changes in pump-blade position are insufficient to obtain the required flow the number of pumps working in parallel can be changed. There is a flow meter in each output pipe; a control circuit keeps the total output evenly distributed over all pumps in opera-

tion. *Fig. 10* shows how the pumps are switched on and off and the output is distributed.

The behaviour of the rapid sand filters — up to 24 of these can be used in parallel — has been described in the previous section. The water level above the sand filter is kept constant, even during variations in the flow, by regulating the output flow. The outflowing filtrate water is stored in underground tanks, whose level is kept constant by means of a control loop.

After the raw treatment the water is transported along a double pipeline to the final treatment station. Here again centrifugal pumps operating at constant speed and with d.c.-actuated blades are used to control the output. Unlike the situation at the input, where any number of identical pumps may be used in parallel, there are in this case three pumps of different capacity and a spare pump equal in capacity to the largest of these. The pump selected is the one most suitable for the desired outflow. The pipe diameters here are greater than at the input. There are flow meters in each pipe; the sum of their readings gives the total output flow.

Hierarchy of the control system

The raw-water treatment plant described above is controlled by means of four control loops: two for controlling the input and output pumps and two for keeping the water level above the filters and in the filtrate basin constant (*fig. 11*). In the control configuration that we have used for the raw-water treatment plant the most important control loop is the one that controls the level of the storage reservoir. Any deviations in the water level are fed back directly to the input pumps. The storage reservoirs in this control loop give a purely integral behaviour, since the level variations are proportional to the integral of the difference between the input and output flow rates. The reference value for the pumps at the output is given directly by the water demand of the final-treatment station.

The variations that the control system must be capable of handling are changes in the reference value for the outflow from the plant (1800-5300 m³/h) and the rapid supply of the water needed for cleaning the filters (3000 m³/h for 5 minutes to clean one filter).

The control concept outlined above, where the pumps at the output are directly controlled by the outflow demand and the pumps at the input are controlled by the signal derived from the storage-reservoir level controller, does not in itself give a completely satisfactory result. For reference-value changes and filter backwashing the variations in the water level are too great and the flow rate through the filters is not sufficiently constant. This situation can be improved by introducing a feedforward type of control, in which the pumps at the input are partly controlled by the flow desired at the output of the raw-water treatment plant. This feedforward control is indicated by an interrupted dashed line in fig. 11.

After introduction of the feedforward control the behaviour of the plant was substantially improved. Deviations from the reference values for the various water levels are now acceptable, and variations in the flow rate through the filters are also considerably reduced. The responses of the various quantities to a change in the reference value of the outflow and to a filter backwashing procedure are shown in *fig. 12*.

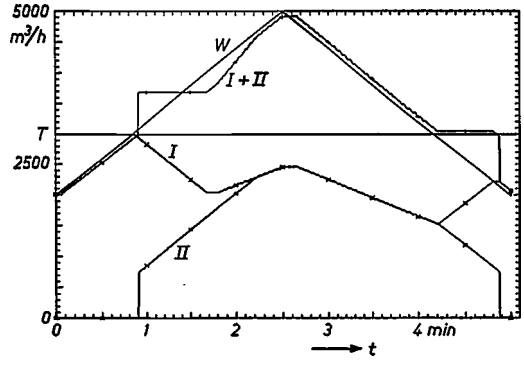


Fig. 10. Manner in which two pumps together deliver the desired outflow *W*, at first increasing with time *t* and then decreasing. *I* and *II* the shares in the total real outflow *I* + *II* delivered by the respective pumps. If the desired value of the outflow is in the neighbourhood of the threshold value *T* for switching over from one to two pumps, continuous switching to and fro is avoided by a slight hysteresis in the control circuit: switching to two pumps takes place at a value *W* slightly above *T*, and back to one pump at a value slightly below *T*.

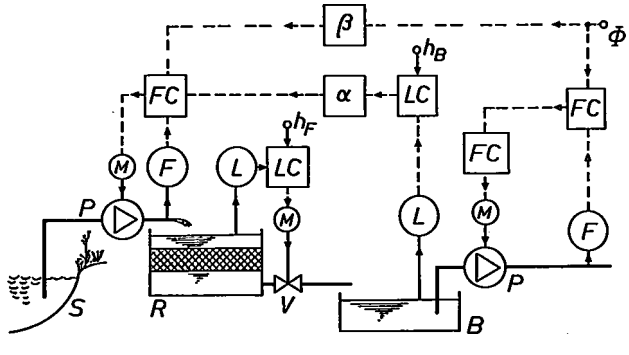


Fig. 11. Diagram of the control of the raw-water treatment plant. *S* storage reservoir. *R* rapid sand filter. *B* storage basin. The heavy lines represent the water pipes, with pumps (*P*) and a control valve (*V*). The thinner lines represent mechanical connections with the motors (*M*) for varying the position of pump blades and valve settings or for the level measurement (*L*) or flow measurement (*F*). The dashed lines are the electrical connections between the various control units (*FC* flow controllers, *LC* level controllers where h_F is the reference value for the water level above the filter and h_B for the level in the storage basin); arrows indicate the direction in which control signals are operative. Φ reference value for the outflow from the raw-water treatment station. This signal may either be used only for controlling the pumps at the output, or also for the feedforward control of the pumps at the input via the interrupted dashed line. The networks α and β determine the degree of feedforward control in this case.

Control of the final stations of the water-treatment plant

Elements of the system

At the input of the final-treatment stations there is a storage reservoir, which has a stand-by function in the event of temporary interruptions in the inflow. The main flow in fact bypasses the reservoir, no more of the inflow passing through it than is required for replenishment. The level of the reservoir is controlled as well as the outflow from it. This control is based on an internal plant reference value and this value also controls the outflow from the raw-water treatment plant. *Fig. 13* shows the behaviour of the reservoir at a change in the value for the inflow. The peaks in the measured inflow are due to switching over to larger pumps at the output of the raw-water treatment plant, necessary for meeting a larger outflow demand.

Next the water flows through an ozonizer station. The level in the output stage of this station is controlled by varying the speed of the pump in the outflow pipe and if necessary by switching on the second pump. The flow signal from the input can be used for switching the pumps. A consequence of the small surface area of the buffer is that the reactions in the water level to a change in the flow rate can be rather marked, as shown in *fig. 14*.

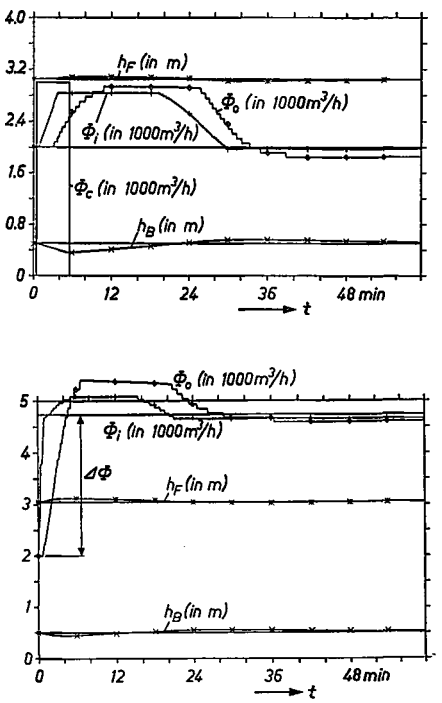


Fig. 12. Behaviour of the raw-water treatment plant after introduction of feedforward control. *Above:* Responses to the backwashing of one of the 24 filters. h_B water level in the storage basin from which the backwashing water is drawn. Φ_i total inflow above the filters. Φ_o total outflow from the filters. h_F water level above the filters in use. *Below:* Response to a change $\Delta\Phi$ in the reference value for the outflow.

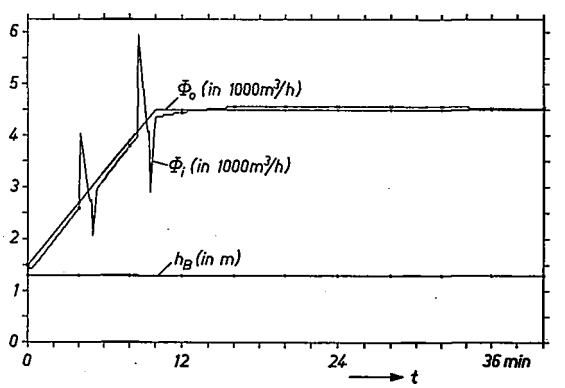


Fig. 13. Response of the storage basin at the input of the final treatment plant to a test function for Φ_o, the outflow demand. h_B water level in the basin. Φ_i inflow. The two peaks in Φ_i are due to switching over twice to pumps of higher capacity at the output of the raw-water treatment plant.

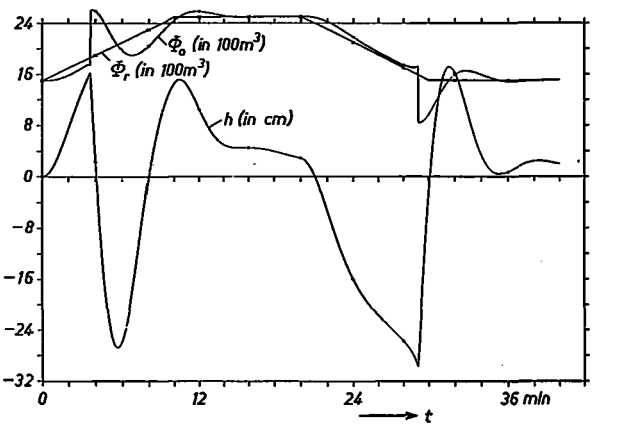


Fig. 14. Behaviour of the ozonizer station to changes in the reference value for the outflow (Φ_r). Owing to the small surface area of the tank, the water level h responds rather strongly to the changes. Φ_o the actual value of the outflow. The peaks are caused by switching to another pump.

The next station in the system, the coagulation basin, is not separately controlled. Under steady-state conditions inflow and outflow are equal; the water level in the basin is determined by heights of a sluice gate and the magnitude of the flow. The system responses and method of controlling the rapid sand filters that now have to be passed have already been described.

In the case of the slow sand filters, which are responsible for the biological cleansing of the now almost completely purified water, both the level above the filter bed and the outflow are controlled. The output signal from the water-level controller acts as the internal reference value for the whole system. It has an effect on the inflow as well as controlling the outflow from the storage reservoir at the input. In the control loop for the outflow from the slow sand filter a delay element is incorporated, ensuring that a change in the flow rate cannot take place more rapidly than the maximum permissible rate of change. Flow meters are again used

here, one in the output pipe of each filter basin and one in the common pipe for the total outflow. The response of the filter to a change in the reference value for the outflow is shown in *fig. 15*.

After leaving the slow sand filters the water is passed through an underground chlorination tank, where chlorine is added to the water to prevent the growth of algae in the mains network. The amount of chlorine added should be such that when the water flows out of the kitchen tap all the free chlorine has just been used up by oxidation of unwanted material in the pipes. A simple level control in the chlorination tank is sufficient to ensure that the outflow is equal to the inflow.

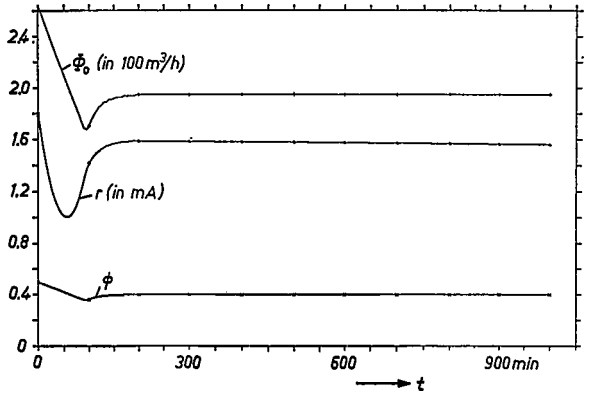


Fig. 15. Responses of the slow sand filter to a change in the reference value for the outflow demand. Φ_0 the outflow that initially decreases at a constant, maximum permissible rate for the slow filter. r output signal of the controller for the total water flow. ϕ normalized valve settings in the outflow pipes of the individual filters.

Control of the treatment plant as a whole

The reference value for the control of the complete water-treatment plant is the outflow demand. Details of the control scheme are summarized in *fig. 16*. The outflow demand directly controls the water flow from the slow sand filters. Changes in this outflow have their effect on the water level above these slow filters. The output signal from the controller for this water level acts as an internal reference value for the final treatment plant, and, owing to the feedforward control in the raw-water treatment plant described above, it also affects the entire plant right up to the pumps at the input of the raw-water station (case *I*). The behaviour of the water-treatment plant controlled in this way may be seen from the reactions of the flow at various points in the plant, shown in *fig. 17*. Some of the reactions are fairly strong, and at the end of the time interval in the figure (200 minutes) the pattern is still not entirely steady. The number of times the pumps have to be switched at the output of the raw-water treatment station appears to be greater than is strictly necessary. This behaviour can be improved by arranging that the reference value for the outflow not only acts on the output of the slow sand filters in the event of changes, but also acts on the internal reference value for the whole plant (case *II*). In this way the desired changes in the outflow are made to exercise feedforward control of the whole plant. One thing that must not be forgotten, however, is that there is a maximum permissible rate of change in the outflow from the slow sand filters.

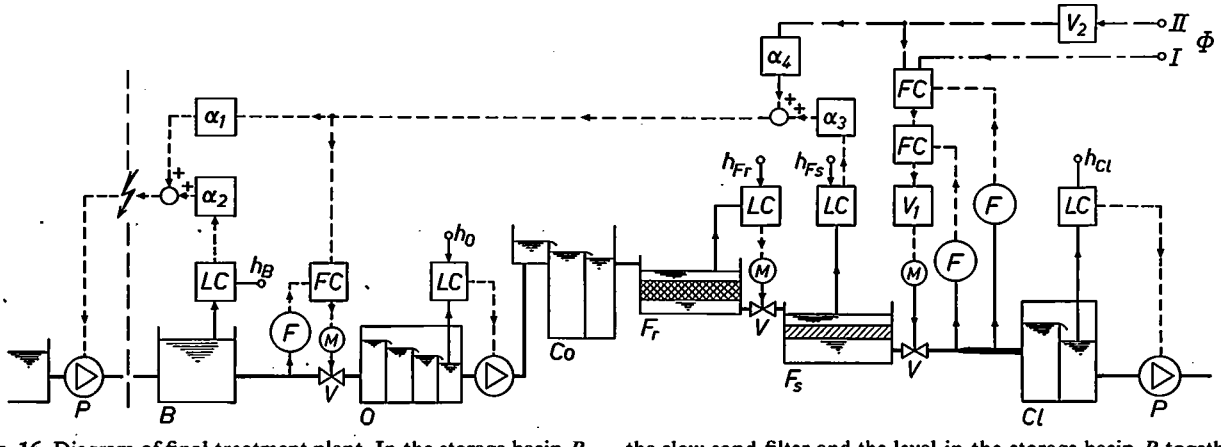


Fig. 16. Diagram of final treatment plant. In the storage basin B, the ozonizer station O, above the rapid filter F_r, the slow filter F_s and in the chlorination station Cl the water level is held at the reference values h by the various level controllers LC, which control the flow from one of the pumps P or the flow through one of the valves V. The coagulation tank Co has no level control. The flow is measured at various points with flow meters F, and kept constant by flow controllers FC which, driven by motors M, operate the valves V. In the output pipeline of the slow sand filters there are two flow-control circuits. One circuit uses the signal Φ for the total flow demand from the plant to control the outflow from the whole filter section. The other circuit ensures that each filter separately delivers a proportional share to the total flow. This circuit controls the valves in the output pipes via a delay circuit V_1, to limit the rate at which the flow changes. The level above the slow sand filter and the level in the storage basin B together determine the flow from the pump at the output of the raw-water treatment station; networks α_1 and α_2 determine the ratio in which these two factors have their effect.

The reference value Φ for the total outflow from the plant can be processed in two ways. In one case (*I*) Φ is used as the reference value of the flow controller for the total outflow only. In the other case (*II*) it is at the same time used as a feedforward control signal for pumps and valves at the beginning of the treatment path. Networks α_3 and α_4 now determine the ratio in which the feedforward control is determined by Φ and by the water level above the slow sand filter. A delay circuit V_2 is also incorporated to ensure that the reference value does not change faster than is permissible for the slow sand filters. Case *II* has been used for the design of the installation.

To avoid excessive variations in the water levels at various places with the proposed feedforward control, it is necessary that changes in the feedforward control signal should also be slowed down to this rate. The behaviour of the system with this form of control is shown in *fig. 18*. After a change in the reference value for the outflow the whole system now returns to a steady state more rapidly. The number of times the pumps are switched at the output of the raw-water pumping station is also reduced to a minimum.

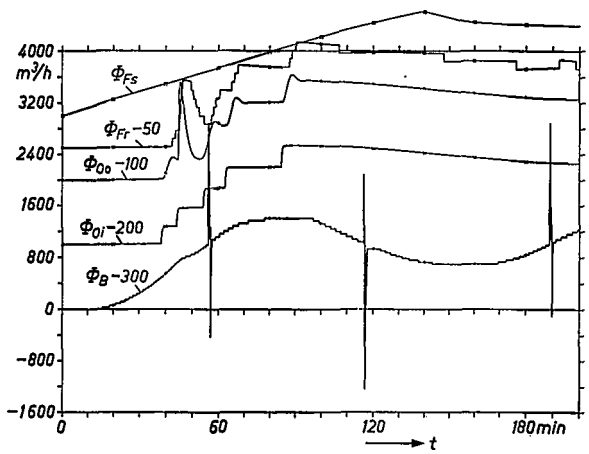


Fig. 17. Responses of the final treatment to a change in the reference value of the outflow when this is controlled directly, (case *I* in fig. 16). Φ_B inflow in the storage basin. Φ_{Oi} inflow in the ozonizer station. Φ_{Oo} outflow from the ozonizer station. Φ_{Fr} outflow from the fast sand filters, Φ_{Fs} outflow from the slow sand filters. The curves are vertically displaced in relation to one another so that they can be displayed in one figure. After 200 min the various quantities have not yet reached their final value.

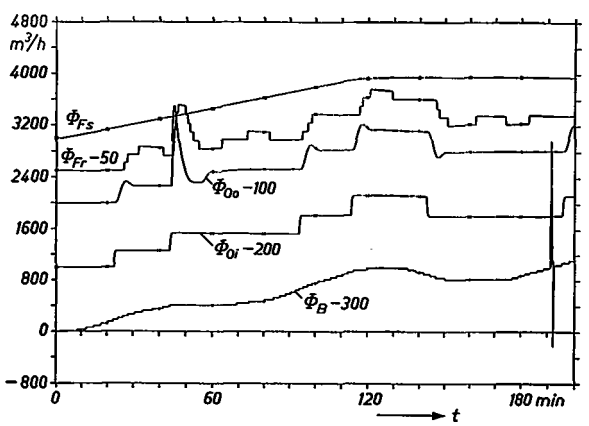


Fig. 18. Responses of the final treatment plant to the same change in the reference value of the outflow as in fig. 17 in the case of feedforward control (*II* in fig. 16). Indications relating to the curves are the same as in fig. 17, but the vertical scale is different. After 200 min all quantities have now reached virtually stable values. It can also be seen from Φ_B that the pumps at the output of the raw-water treatment plant have now been switched only once, rather than three times as in fig. 17.

The feedforward control system does however require extra vigilance in abnormal situations. If, for example, the outflow pipe from the storage basin has to be shut off for one reason or another, then in the first case with the direct control the rise in the water level will slow down the pumps in the raw-water pumping station. With feedforward control, however, only a fraction of the pumping-station reference value is due to the water-level control; the remainder stays constant in spite of the rising level. In such cases a special warning signal must be given so that the operators in the central control room can take action to prevent overflow in the storage reservoir.

Notwithstanding this minor drawback, the results obtained with feedforward control may be described as satisfactory. This type of system, the details of which have been obtained by computer simulation, will therefore be used for the control of the water-treatment plant. The complete system will consist of electronic controllers. A computer is installed for the regular compiling of data concerning the operation of the plant and for determining deviations from the normal state.

In the design phase of the control system much help was received from Ing. B. van Zutphen and Ir W. L. Prinsen Geerligs and their staff of the Amsterdam Water-Supply Authority, and from Ing. H. A. Verharen of the Philips Scientific and Industrial Equipment Division. They provided detailed information on the purification plant and valuable help in evaluating the simulation results.

Summary. The drinking water for the Amsterdam Water-Supply Authority is obtained from a complex plant in which surface water is treated by a system of filters and ozonizer and coagulation stations. This article describes the design of an automatic control system for this plant, to prevent pumps from running dry and reservoirs from overflowing and to ensure that the filters function optimally and do not run dry. From the various possible ways of designing a control system of this type, a simulation-aided design method was chosen, using a hybrid computer, i.e. one that is partly analog and partly digital. From the result of these simulations it is possible to determine how and to what extent the feedforward approach can be used to obtain the desired behaviour from the complete system.

An experimental system for the automatic determination of organochlorine residues in milk

R. J. Dolphin, L. P. J. Hoogeveen and F. W. Willmott

During the last thirty years, pesticides have been used on an increasingly large scale in agriculture with a view to stepping up food production to meet the world's ever-growing food requirements. Since organochlorine pesticides are a potential health hazard, many countries have strict legal regulations requiring analyses of certain foods. A need has therefore arisen for equipment that can carry out such analyses quickly and accurately. The article below describes a system that automatically monitors organochlorine residues in milk.

Introduction

The chemical agents used to control pests in agriculture (pesticides) may find their way into our food, either unmodified or metabolized by plants or animals. Since these substances ('residues') above a certain concentration are a possible hazard to human health, it is desirable to subject foodstuffs to regular analyses. In many countries this is strictly regulated by law. It is particularly important to analyse the organochlorine pesticides and their metabolites that accumulate in fat, such as DDT, DDE, lindane, hexachlorobenzene and the polychlorinated biphenyls (PCBs). A major food that contains fat is milk. To measure the ten residues most often encountered in milk, at a concentration of the order of 10^{-9} g/l, it is desirable to have a monitor that is capable of carrying out rapid routine analyses.

In this article we describe an experimental monitor developed at two of the Philips research laboratories (Philips Research Laboratories, Eindhoven, and Mullard Research Laboratories, Redhill), which is the first in the world to offer the possibility of automatic analysis. The monitor can process two samples an hour, and in each analysis it determines the fat content in addition to the residues of the ten named substances.

Design basis; general features of the monitor

Before describing the system in more detail, we shall begin by looking at some of the factors that proved to be important during its development.

Dr R. J. Dolphin and Dr F. W. Willmott are with Mullard Research Laboratories, Redhill, Surrey, England; Drs L. P. J. Hoogeveen is with Philips Research Laboratories, Eindhoven.

After the sample has been taken, the analysis can generally be divided into three distinct operations: pretreatment of the sample, separation into individual components and measurement (detection).

The first operation is to make the sample suitable for further processing; this includes the removal of undesirable constituents from the sample and bringing the sample to an appropriate concentration for the separation. In the next stage the residues to be detected have to be separated to allow the amount of each constituent to be detected with sufficient accuracy. The third stage is the actual measurement. Here it is not only the accuracy of the measurement that is important, but also the detection limit that can be reached, which must be well below the value considered as the maximum permissible limit.

Residue separation is normally carried out by means of chromatography, which has been described as 'an analytical technique for the resolution of solutes, in which separation is produced by differential migration in a porous medium, and migration is caused by flow of liquids or gases' [1].

In the most common form of this separation method, column chromatography, the sample to be analysed is first injected into a narrow zone at the end of a tube filled with tightly packed absorbent granular material (the column). Next, the column is flushed through with a liquid (in liquid chromatography) or with gas (in gas chromatography). This flushing process, called elution, causes a displacement of the soluble fractions of the various constituents of the injected sample. The physical separation arises because the displacement

does not take place at the same rate for all substances. The difference in the rate of displacement may be due to differences in the adsorption on the grains in the column (adsorption chromatography) or to differences in the distribution coefficient, for instance when a stationary liquid layer is present around the grains (partition chromatography). The interaction of the various constituents of the sample with the eluent and with the packing in the column determines the eluent

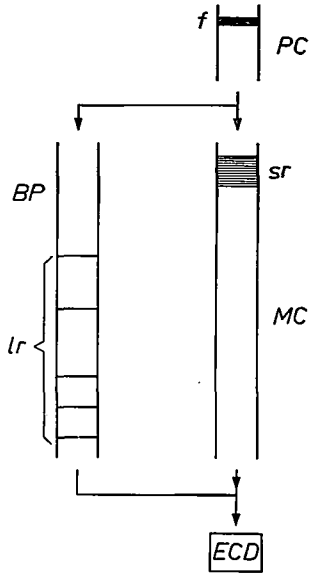


Fig. 1. Diagram showing the way in which the two separation columns are used in the automatic milk pesticide monitor. During elution the residues pass through the short precolum *PC* and the fat *f* is retained. The residues *sr* with a short retention time leave this column first and are temporarily stored in the long main column *MC*; at this stage they are not yet sufficiently separated. The residues *lr* passing later, which have a long retention time, emerge from the precolumn completely separated and are led through the bypass coil *BP* to the electron-capture detector *ECD*. Once these residues have been detected and measured, the first group *sr* is then subjected to further elution to complete the separation, and the components of this group are then detected and measured.

and flow that should be selected to obtain the desired separation. The elution also serves to transport the different constituents to the other end of the column, where they are separately available for detection. The time necessary to transport a substance from one end of the column to the other is characteristic of the substance and is called the retention time.

Gas chromatography has previously been the method used for the analysis of residues in milk, largely because of the very low detection limit obtained with the detectors used. For reasons that will presently be discussed, we have chosen *liquid* chromatography for the separation. To give rapid analysis, two separation columns are used.

Before looking more closely at the design of the separation stage, we shall say a few words about the two other stages: the pretreatment and detection stages. The pretreatment can be fully automated by using the droplet or 'bubble-train' extractor [2], developed at Philips Research Laboratories. In this device the fat with the residues contained in it is extracted from milk with n-hexane. The same liquid is also used as the eluent in the liquid chromatography. The third stage, in which the constituents are measured, includes a new application of electron-capture detection. With this detector the required detection limits can still be achieved in a monitor in which the separation is carried out by means of liquid chromatography.

The introduction of the three new features mentioned above — extraction with a bubble-train extractor, separation by means of liquid chromatography and measurement with a new type of detector — has made it possible to achieve full automation and a considerable reduction in analysis time [3].

Liquid chromatography with two columns

To give rapid separation two columns are used: a short one, the 'precolumn' (5 cm), and a long one, the 'main column' (15 cm); *see fig. 1*. The primary purpose of the precolumn is to separate the residue from the fat; during the elution the residues move through the precolumn much faster than the fat, which remains behind.

After each analysis it is of course necessary to clean the columns and the other parts of the apparatus completely before the next sample is injected. The main problem in this 'regeneration', as the cleaning process is called, is the excess fat present in the column. The fat can only be removed sufficiently rapidly by using a second, more polar eluent, which in our case is n-hexane with 1% of isopropylalcohol added. A disadvantage of using this eluent, however, is that even small traces of the alcohol considerably reduce the resolution of the columns. For this reason the column has to be flushed again with pure n-hexane. If only one long separation column were used, the regeneration would therefore be too time-consuming. With our two-column method, since all the fat is retained in the short precolumn, the main column requires no regeneration and the cycle time is reduced.

Elution through the precolumn gives a separation not only between the fat and the residues but also between groups of residues. Some residues move through

[1] See for example B. L. Karger, L. R. Snyder and C. Horvath, An introduction to separation science, Wiley, New York 1973.
[2] A. van Tooren, Philips tech. Rev. 35, 196, 1975.
[3] A more detailed description will be found in R. J. Dolphin, F. W. Willmott, A. D. Mills and L. P. J. Hoogeveen, J. Chromatogr. 122, 259, 1976, and in L. P. J. Hoogeveen, F. W. Willmott and R. J. Dolphin, Z. anal. Chemie 282, 401, 1976.

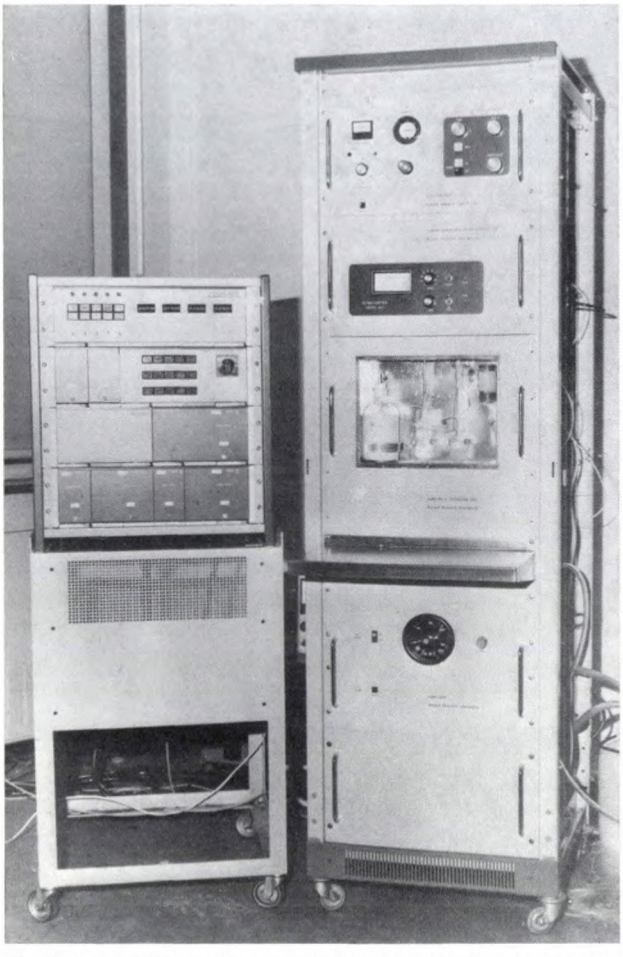


Fig. 2. View of the milk pesticide monitor, clearly showing the modular structure. The cabinet on the right contains, from top to bottom, racks for the measuring instruments, the columns for the chromatographic separation, the pretreatment unit, and the service unit containing the pumps and the store of chemicals. The unit on the left is a microcomputer, which controls the pumps and the various valves, supervises the stocks, calculates the results and gives the alarm in the event of malfunctioning.

rapidly (short retention time) and are barely separated from each other when they leave the column, but there is also a group that moves slowly through the precolumn and in the process is separated into its components. In the system we have designed, the flow of eluent containing the second group of residues is led through a 'bypass coil' (fig. 1) to the detector after the first group has been eluted from the precolumn and has reached the main column. The first group is stored in the main column until the detection of this second group has been completed. After the second group it is then the turn of the first group for further elution and detection.

In this way the group with the longest retention times is passed only through the short precolumn, which again saves time.

The small movement of the fat in the precolumn during the separation of the residues led to the idea of eluting the fat in reverse (back-flushing). This gives a further saving in time.

Design and operation of the monitor

A photograph of our experimental 'milk pesticide monitor' is shown in *fig. 2*. The modular structure can clearly be seen. The rack at the upper right contains the detector unit, the rack below it the separation module; below it are the module for the pretreatment of the sample and the 'service' module, which contains the pumps and the store of chemicals. On the left of the photograph is the microcomputer, whose function is to control the pumps, the detector, the valves etc., to calculate and report the results, and to give the alarm in the event of malfunctioning of the equipment or if the contents determined exceed critical values [*].

The construction and operation of the monitor will now be described in more detail with reference to the diagram in *fig. 3*.

Pretreatment

A 25-ml sample *S* is taken from the milk to be analysed and is thoroughly mixed in *M* with 25 ml of dilute sulphuric acid from *A*. The sulphuric acid removes the adsorbed proteins that keep the milk fat in emulsified droplets and the fat is thus liberated. In the next stage the liberated fat is extracted with n-hexane in the extractor *E*, shown separately in *fig. 4*.

In the last stage of the pretreatment the hexane fraction is passed over a drying column *D*, which first breaks down any remaining emulsion, and then removes residual sulphuric acid and water from the hexane fraction. While the sample and later the hexane fraction are going through the various stages of the pretreatment, the components of the equipment that have fulfilled their function are flushed clean.

The complete pretreatment, including flushing for the next analysis, lasts half an hour. The extract then goes to the separation module, where it again remains for half an hour. During this time the next sample can be pretreated.

The sequence of events in the extractor is as follows. The peristaltic pump P_3 pumps the n-hexane and the mixture of milk and sulphuric acid through a Y-type mixing piece *Y* to produce a row of alternate droplets of these two liquids, the 'bubble-train', in a tube *T* five metres long. The extraction takes place in the tube. The bubble train is then fed into a glass separating vessel *Sep*, where the n-hexane is separated from the milk/acid mixture. The two fractions are then pumped back to the Y-piece and the process is repeated. After it has circulated about twenty times through the extraction tube, the n-hexane has extracted the fat and the residues dissolved in it with an efficiency of about 50 %. The extraction need not be complete, or be carried out with an accurately defined efficiency, since the analysis of both the fat and the residues is performed on the

same extract, and the residues are contained mainly in the fat. In spite of variations in the extraction efficiency, the amount of residue can therefore always be indicated accurately as a fraction of the quantity of fat.

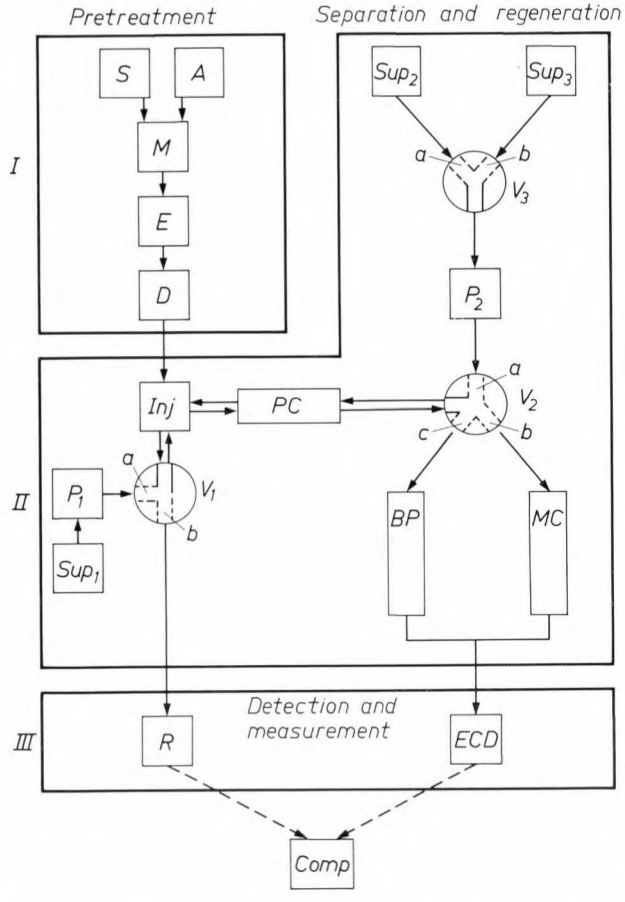


Fig. 3. Diagram of the milk pesticide monitor. Blocks *I*, *II* and *III* are the pretreatment unit, the separation and regeneration unit, and the detection unit respectively. (In reality part of the separation and regeneration equipment, pumps, etc. are contained in a separate service module.) *S* and *A* are the containers for the sample and the acid (dilute sulphuric acid). *M* mixer. *E* bubble-train extractor (see fig. 4). *D* drying column. *Inj* injection valve. V_1, V_2 and V_3 valves. P_1 and P_2 pumps. Sup_1, Sup_2 and Sup_3 supply vessels containing respectively n-hexane, n-hexane plus 1% of isopropylalcohol, and n-hexane. *PC* precolumn, 5 cm long. *MC* main column, 15 cm long. *BP* bypass coil. *ECD* electron-capture detector for determining the residue contents; see also fig. 6. *R* differential refractometer for determining the fat content. *Comp* microcomputer.

Separation and regeneration

The separation is carried out by means of a modern form of liquid chromatography, known as high-performance liquid chromatography [4]. For this method both the precolumn *PC* and the main column *MC* are packed with microparticulate silica (particle diameter 10 μm). Two pumps P_1 and P_2 cause the various eluents from the supply vessels Sup_1, Sup_2, and Sup_3 to flow in the desired direction through the columns under a pressure ranging from 6.10^6 to 9.10^6 Pa (60 to 90 atm).

The microcomputer *Comp* contains a program that ensures that the positions of the various valves are changed at the right moment. The procedure is then as follows.

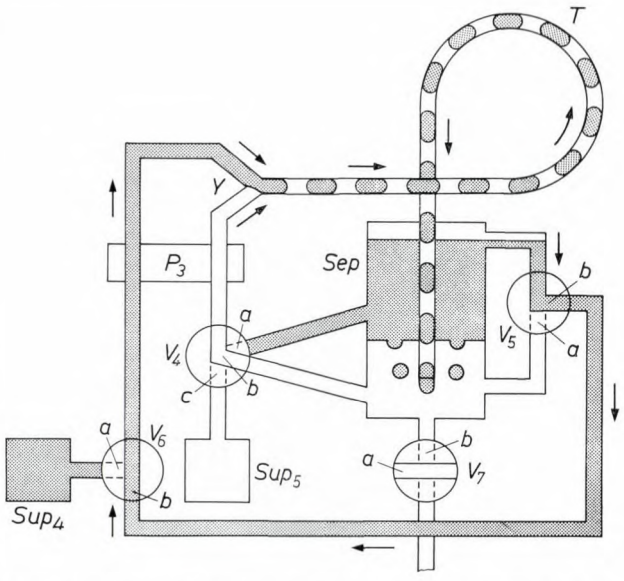


Fig. 4. Diagram of the extractor used in the monitor. P_3 peristaltic pump. *Y* mixing piece. *T* extraction tube. *Sep* separation vessel. V_4, V_5, V_6 and V_7 valves. Sup_4 and Sup_5 supply reservoirs containing respectively the extraction medium n-hexane and the pretreated milk.

In the separation module (*II*) 100 μl of the extract is injected into the precolumn by means of an injection valve *Inj*. Details of this part of the module can be seen in *fig. 5*. The pump P_1 directs the eluent n-hexane from Sup_1 through valve V_1 in position *a* and V_2 in position *b* through the precolumn and the main column, until all components with short retention times are held in the main column. Next, valve V_2 is set to position *c* so that the elution flow is directed through the bypass coil *BP* to the electron-capture detector *ECD*. Since this detector is sensitive to pressure fluctuations, the bypass coil is dimensioned in such a way that the pressure drop across it is the same as that across the main column. Once all the residues supplied through *BP* have been detected and measured, valve V_2 is reset to position *b* so that the eluent is fed through the main column to the detector. After the residues temporarily stored in the main column have also been separated and measured, pump P_1 is switched off and pump P_2 comes into operation. With the aid of this pump and a mixture

[4] See for example L. R. Snyder and J. J. Kirkland, An introduction to modern liquid chromatography, Wiley, New York 1974.

[*] Valuable help in the systems analysis and in the design of the electronic equipment was given by Mr L. Tober of the Philips Scientific and Industrial Equipment Division (S & I), and Mr A. W. van der Meulen of Philips Research Laboratories, Eindhoven.

of n-hexane and 1% of isopropylalcohol from Sup_2 (V_3 in position *a*, V_2 in position *a*, and V_1 in position *b*) the fat from the precolumn is eluted and led to the differential refractometer *R*, which determines the fat content. After V_3 has been set to position *b*, the precolumn is next regenerated with n-hexane from Sup_3, while valves V_3, V_2 and V_1 and the injection valve *Inj* are flushed to prevent contamination of the next sample.

contained in the vapour, which capture these secondary electrons; this current reduction is therefore a measure of the quantity of residue molecules present. The identification of the residues is performed by means of a calibration based on retention times.

The hexane vapour is expelled from the detector by passing a constant flow of nitrogen gas through it to a condenser.

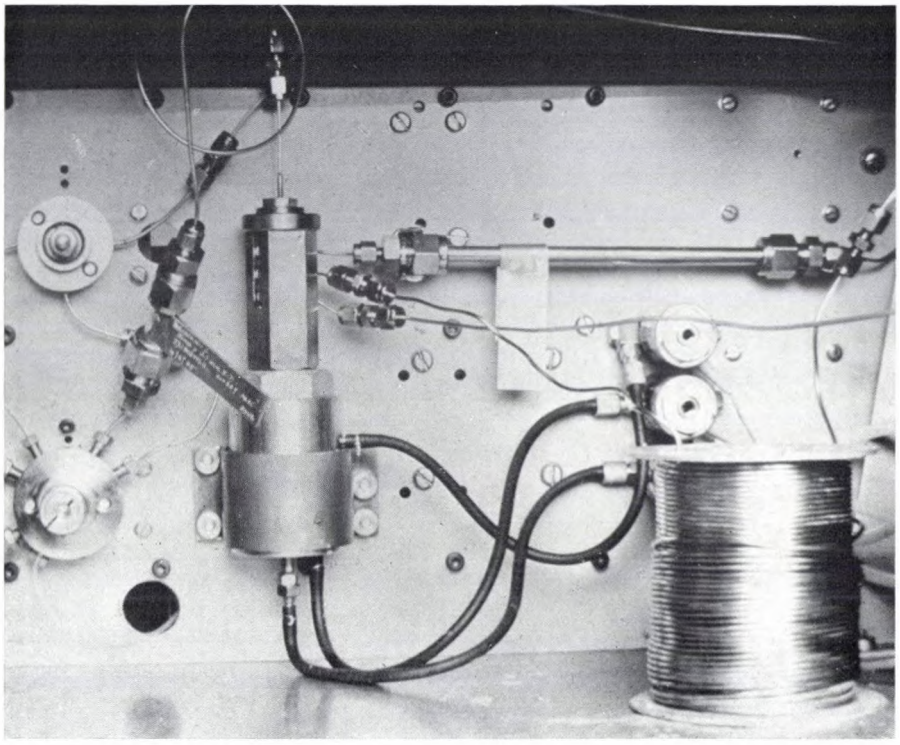


Fig. 5. Inside view of the separation module. The vertical component, to the left of the centre, is the electromagnetically controlled valve V_2 (fig. 3). This valve directs the eluent either to the main column *MC* (horizontal tube on the right) or to the bypass coil *BP* (lower right), or to the obliquely situated precolumn *PC* to the left of it. The injection valve *Inj* is at the lower left and the valve V_1 is above it.

Measurement

The system we use for the detection of the organochlorine residues (*fig. 6*) is based on the same principle as that of the electron-capture detectors used in gas chromatography, but is of course adapted to the use of liquids. The eluent from the separation columns is passed through a heated steel capillary. The eluent evaporates and the resultant pressure increase causes the vapour to be directed through the detector. Beta radiation, emitted by the radioactive nucleide ^{63}Ni, ionizes the molecules of n-hexane in the eluting vapour, and generates a current of secondary electrons between an electrode held at a positive potential and the housing of the detector. This zero current or 'standing current' is reduced by the molecules of organochlorine residues

Results

In *fig. 7* the separation using our two-column method is compared with that using a single column. Fig. 7*a* shows the time required for adequately separating all the indicated residues with one column packed with the same material as in our case and operating under otherwise identical conditions. The separation takes about 30 minutes. Fig. 7*b* gives the time required in our method for elution through the precolumn; only the group with the long retention times, beginning with γ-BHC, is then adequately separated. Fig. 7*c* gives the total time required in our method for separating all the residues, eluting the fat (T_4) and regenerating the precolumn (T_5). In any further comparison of the two methods it is necessary to remember that the last two

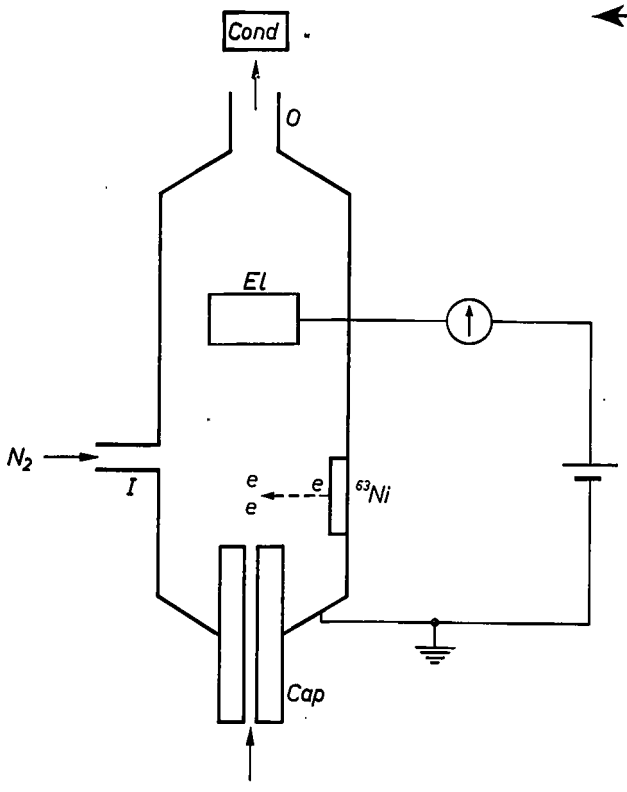
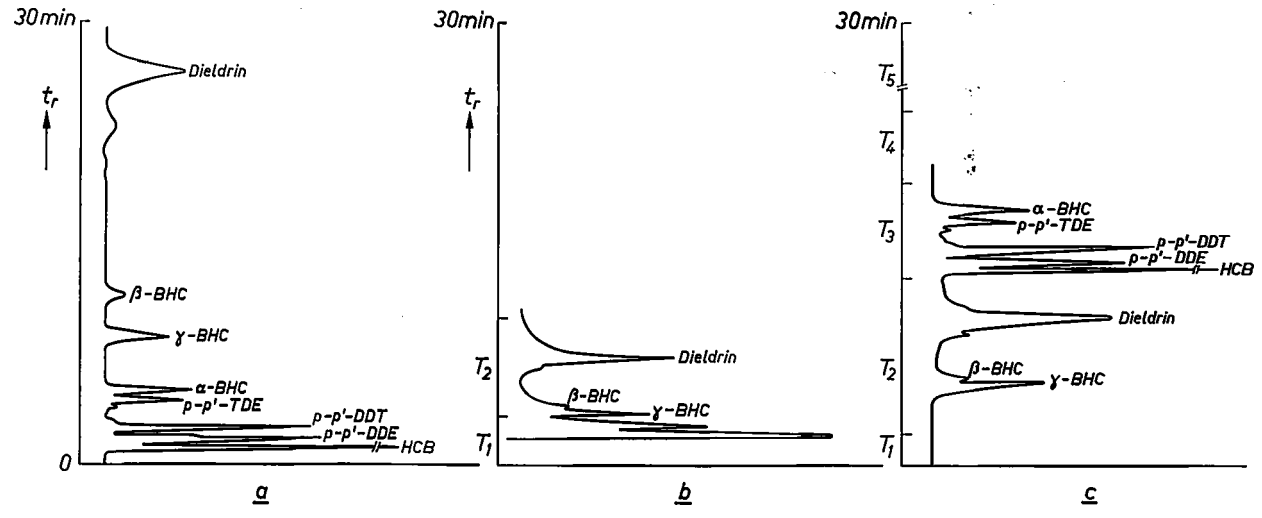


Fig. 6. Diagram of the electron-capture detector used for liquid chromatography. The n-hexane from the separation columns flows through a heated stainless-steel capillary *Cap*, where it is evaporated so that it enters the detector as a vapour. Beta radiation, emitted from the radioactive nucleide ^{63}Ni, ionizes the incoming hexane vapour, producing secondary electrons that are conducted to the ring-shaped electrode *El*, which is held at a positive potential. The resulting standing current is reduced by electron-capturing residue molecules, and the reduction is a measure of the quantity of these residues. A continuous stream of nitrogen gas is passed through *I* to *O* to expel the residual vapour to the condensor *Cond*.

Fig. 7. The degree of separation and the time it takes with a single-column method and with our two-column method, under comparable conditions. The graph (*a*) indicates the retention time t_r that the various residues take to pass through a single column sufficiently long for all the residues to be completely separated, and graph (*b*) shows the retention times taken to pass through our precolumn. In passing through our precolumn some of the residues are not completely separated (times T_1) whereas some are completely separated (times T_2). The graph (*c*) gives the total time required for the analysis: T_3 is required for the further separation in the main column of the residues with a short retention time, T_4 is the time required for eluting the fat, and T_5 the time required for regeneration of the short precolumn. The total analysis time with our method is 30 min. If (*a*) included the time required for the same processes in the single-column method, an estimated 1½ hours would have to be added to the half hour required for the separation alone.



processes would require an estimated additional time of at least 1½ hours if the single-column method were used, which would make the total analysis time 2½ hours.

Table I lists the maximum permissible content of ten residues in milk as quoted by the Netherlands Ministry of Agriculture. The requirements that the monitor should satisfy can be derived from these figures; in particular the detection limit for each separate compound. The table also gives the detection limits so far achieved with our system.

The detection limits of our system are not as low as those of the electron-capture detector used in gas chromatography. This is partially explained by a two-fold increase in noise, but the major contribution is due to the background of electron-capturing impurities in the eluent vapour, which reduces the probability of electron capture by specific molecules of interest. This disadvantage is largely compensated by the much larger injection volume permitted in liquid chromatography. Calculations we have carried out [5] indicate not only that the injected volume can be as much as 100 μl without any significant loss of accuracy or resolution (this is at least twenty times the quantity used in gas chromatography), but also that, if a few practical provisions are

[5] R. J. Dolphin and F. W. Willmott, J. chromatogr. Sci. 14, 584, 1976.

Table I. *MAC*: maximum admissible content of ten residues in milk, as specified by the Netherlands Ministry of Agriculture. *MDC*: minimum detection limits laid down by the Ministry. *ADL*: actual detection limits in our monitor. (The Ministry expresses the admissible content and the specified detection limits in mg of residue/kg fat. We have converted these values into g residue/l milk, assuming that milk contains an average of 4% of fat, that a milk sample of 25 cm^3 is taken and that 10 cm^3 of n-hexane is used as extraction medium with an extraction efficiency of 50%.)

Residue	*MAC* (10^{-9} g/l)	*MDC* (10^{-9} g/l)	*ADL* (10^{-9} g/l)
HCB	1.5	0.25	?
α-BHC (α-HCH)	1.5	0.25	0.5
β-BHC (β-HCH)		0.25	0.5
γ-BHC (γ-HCH)		0.25	0.5
Dieldrin	0.75	0.25	0.5
Heptachlor epoxide	0.75	0.25	0.5
p,p'-DDT	6.25	1.5	0.5
p,p'-DDE		1.5	0.5
p,p'-TDE (-DDD)		1.5	0.5
PCBs	2.5	1.5	0.5

made, this volume can be increased still further to achieve the desired detection limits.

Table II shows the reproducibility of the measurements with our monitor, expressed in the small variation of the retention times found.

Although the choice of separation by means of liquid chromatography may have made it slightly more difficult to reach the required detection limits, two analytical advantages, together with the proved possibility of automation and rapid analysis, have decided our choice.

The first advantage is that the separation of DDT from the PCBs, which has always been difficult with gas chromatography, presents no problem with our method.

The second advantage is related to the fact that in conventional methods the sample has to be subjected to several operations during the pretreatment that cause considerable dilution. Since such dilutions would make the method practically unusable by effectively degrading the sensitivity, each of these operations has to be followed by concentration by evaporation of the solvent. Such processes are not only difficult to automate, but also reduce the accuracy of the analysis, since certain volatile residues partly evaporate with the vaporizing solvent and are thus lost for the analysis. In our method, concentration during the pretreatment is maintained by means of the extractor, and evaporation in this stage of the analysis is not necessary. Evaporation is carried out in a closed system immediately prior to the detector, no residues are lost and the accuracy of the analysis is not reduced.

Table II. The absolute retention time t_r and the associated standard deviation σ of nine residues during elution in our two-column system.

Residue	t_r (s)	σ (s)
γ-BHC	267	5
β-BHC	289	7
Heptachlor epoxide	325	9
Dieldrin	610	19
HCB	754	4
p,p'-DDE	766	4
p,p'-DDT	825	5
p,p'-TDE	962	5
α-BHC	1004	4

Summary. An experimental monitor is described that automatically determines ten organochlorine residues in milk. A modular construction has been adopted and each module has its own function: the pretreatment of a sample, the separation into distinct constituents and 'regeneration' of the apparatus, detection, etc. A microcomputer performs the required calculations and has a number of control functions. The monitor uses a bubble-train extractor, high-performance liquid chromatography with two separation columns and a new application of electron-capture detection. Rapid analyses can be performed (two samples an hour) permitting accurate determination, with low detection limits, of the specified residues and the fat content of the milk.

Switching telephone and video-telephone signals

J. W. Coenders

One of the many things that have changed in shape and form under the influence of modern developments in electrical technology is the automatic telephone exchange. These changes came at an appropriate moment; the rapid growth in the number of telephone connections and the increase in the cost of maintenance made it necessary to replace the complicated electromechanical switches such as rotary selectors, which require so much room, by something else. 'Reed relays', which are small, have come into service, and semiconductor elements are also used as switching elements where possible. The control is provided by a 'processor', which can be programmed as required. An example of a completely electronic telephone exchange is the experimental video-telephone exchange at Philips Research Laboratories in Eindhoven, which switches signals with a bandwidth of 1 MHz. Even this exchange itself has already become to some extent old-fashioned; in the meantime it has been found possible to successfully integrate P-N-P-N elements, serving as crosspoint switches, in groups, which considerably simplifies the wiring. The article gives a general account of recent developments.

Since electronics, and in particular integrated circuits, first came into existence the number of different types of telephone exchange has increased markedly. Great diversity exists, especially in switching devices and switching methods, and to a lesser extent in the control because all modern exchanges are controlled by a processor (a digital computer). And while there is still considerable activity in control methods, because the search continues for systems that are as economic as possible and at the same time sufficiently reliable, and also because the newly available adaptability has not as yet been fully exploited for providing new subscriber facilities, the main breakthrough, from wired-logic control to stored-program control, has already taken place.

Besides the large rotary selectors (*fig. 1*) we now find smaller switches in which the contacts are arranged in matrices and formed, for example, by reed relays (*fig. 2*). Electronic devices are as yet used only sparingly. It is doubtful whether any one particular type of switch will predominate in the future, for new forms of communication such as data traffic and video telephony are creating their own special requirements.

In this article the various switching devices and switching methods are reviewed and arranged in 'tree'

Ir J. W. Coenders is with Philips Research Laboratories, Eindhoven.

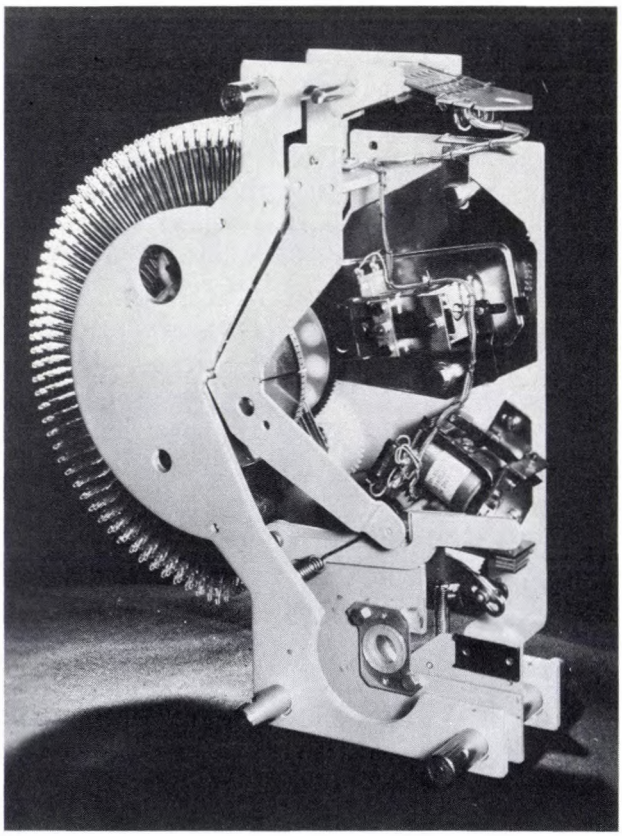


Fig. 1. An example of a conventional electromechanical switch (rotary selector) in a telephone exchange.

form, and the technical arguments leading to a particular choice are examined. Switching operations employing *P-N-P-N* semiconductor elements are then considered at greater length; these are a feature of the experimental video-telephone exchange at Philips Research Laboratories, Eindhoven. The article concludes with a discussion of this exchange. The impedance ratios in this exchange are designed in such a way as to reduce to a minimum the crosstalk between the wideband signals ('high-low principle').

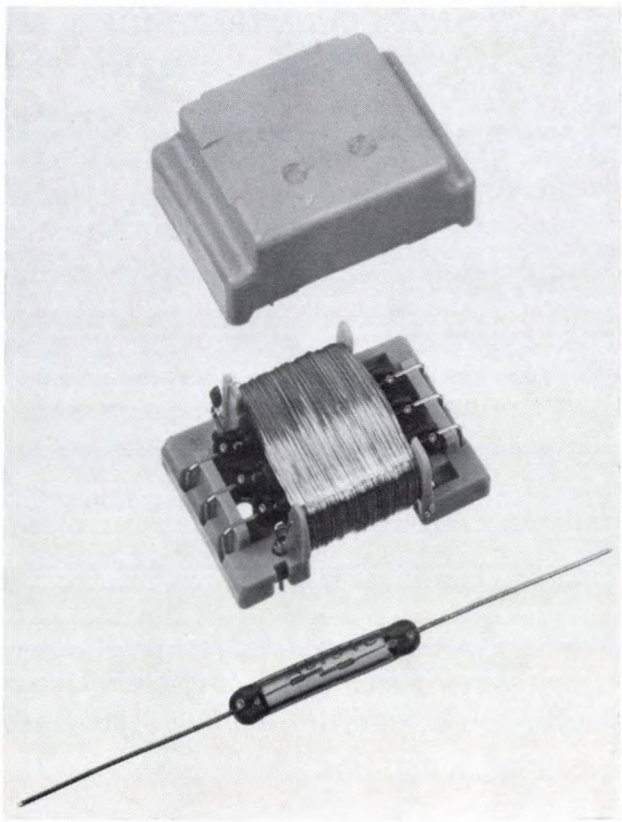


Fig. 2. A reed relay with its energizing magnet coil. The contacts are located inside a sealed glass envelope containing a non-oxidizing atmosphere and are actuated by the magnetic field. Reed relays have taken over the role of rotary selectors (fig. 1) as electromechanical switches in more modern exchanges.

Switching devices and methods

The design of a switching system for telecommunication links necessitates a number of fundamental decisions. These relate to the choice of the switching devices and the switching method, the transfer of control information between subscriber and exchange and between individual exchanges (signalling) and also the way in which the control is organized. In the switching devices and switching methods the most important decisions are the following:

— The switch: electromechanical or electronic. Representatives of the *electromechanical* kind are selectors and relays. One widely used relay is the reed relay. The *electronic* switches may include semiconductor elements such as the diode, the bipolar transistor, the MOS and other field-effect transistors, the four-layer device (*P-N-P-N*), and also include the gas tube.

— The mode of operation of the switch: space-division multiplex (SDM) or time-division multiplex (TDM). In SDM a switch is assigned to a single connection during the entire duration of that connection. This is the situation in conventional exchanges. In TDM a switch makes connection periodically for a short time to transmit signal samples. The same signal path can be used during the intervening time intervals for many other connections.

— The form of the signal: analog or digital. A signal is said to be analog if the information-carrying characteristic of the signal (e.g. amplitude, phase) is continuously variable between certain limits. Signals are said to be digital if only a restricted number of values of the information-carrying characteristic are defined. When digital signals are used, especially if they are binary signals, many advantages are available because of the recent rapid developments in digital circuits; existing modules such as gates and memory elements can be employed without modification. Since, however, much information is available in analog-signal form and is processed in that form, analog-to-digital conversion cannot be avoided.

— The signal transmission: four-wire or two-wire. If a conductor pair is used to transmit signals in only one direction, a circuit intended for transmitting conversation will require two conductor pairs and in this case we talk of a four-wire circuit. If signals are transmitted on a single conductor pair in both directions, the arrangement is known as a two-wire circuit. If the signal voltage on one conductor in a pair is at all times equal and opposite to that on the other, then we have balanced transmission. This is usually limited to transmission over fairly large distances. Within an exchange, however, transmission is sometimes unbalanced, which means that one of the two conductors is used in more than one circuit. This is frequently the case in electronic exchanges.

The alternatives quoted above can be arranged to form a 'tree' of possible choices, with branches leading to a large number of different exchanges [1] (*fig. 3*). Not all combinations of possibilities are meaningful. Types *A* and *B* include all conventional exchanges with electromechanical switches but also modern systems

[1] T. H. Flowers, Electronic crosspoints for switching analogue transmission circuits, Proc. IEE **117**, 65-71, 1970.

[2] P. J. May and T. M. Stump, Synthesis of a resonant transfer filter as applied to a time division multiplex system, Trans. AIEE I **79**, 615-620, 1960.

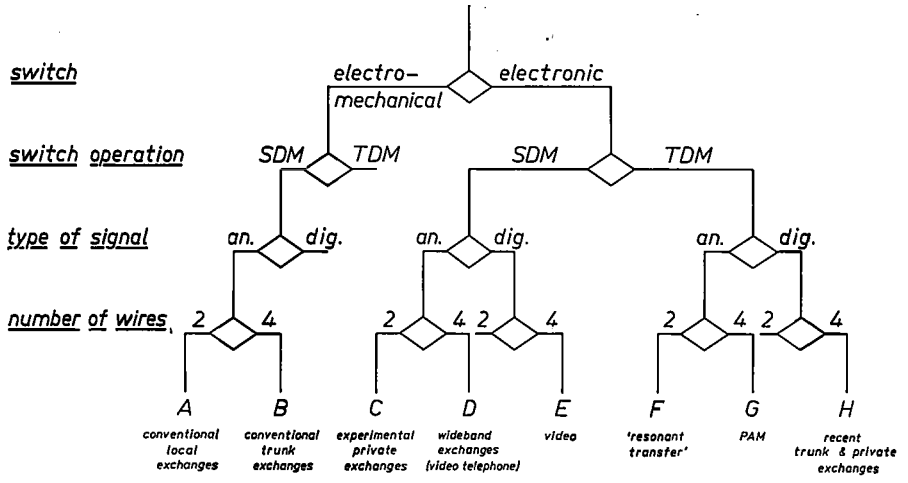


Fig. 3. The various types of telephone exchanges arranged in the form of a 'tree' of choices, showing the type of switch (electromechanical or electronic), how the switches are operated (space-division multiplex (*SDM*) or time-division multiplex (*TDM*)), the form of the signal (analog or digital) and the number of wires for a two-way circuit (two or four). Not all combinations are of significance.

with processor control such as the Philips PRX (processor-controlled reed exchange). Type *C* includes numerous experimental systems that are mainly equipped with transistor and *P-N-P-N* switches. One example of type *D* is the video-telephone exchange at Philips Research Laboratories. Category *E* includes switching systems for digitally encoded video signals. The principle of resonant transfer, now hardly ever used, can be included under *F* [2]. Pulse-amplitude modulation (PAM) is however used today in medium-sized private branch exchanges (category *G*) because the switching devices for TDM are cheaper than the conventional switches.

Digital TDM (category *H*) is of great importance for the future. Its principle can be explained with reference to *fig. 4*. We shall assume that the information transfer between subscribers and the exchange takes place in the form of binary signals. The bits arriving at the exchange from subscriber x are stored in the input buffer B_{ix} and read out periodically at an increased rate, arriving in the output buffer B_{oy} for the desired subscriber y via switch S_1 and the correctly positioned switch S_2. From B_{oy} the information is transmitted to subscriber y at the original bit rate. One of the switches S_1 and S_2 can scan contacts $1 \ldots n$ cyclically, while the other is controlled via a memory containing information about all the desired connections.

Problems in the introduction of electronic devices

It may seem puzzling that the choice made from the possibilities presented by the 'tree' in fig. 3 is not invariably from type *H*. The explanation is to be found in the practical requirements, which are often difficult to meet economically with electronic devices. These requirements are related to factors such as the voltages and currents that have to be transmitted, to the signal attenuation and to crosstalk.

The voltages commonly used on subscribers' lines (a d.c. voltage of 48 to 60 V for the power supply, plus an r.m.s. a.c. voltage of 90 V at 25 Hz for ringing) and

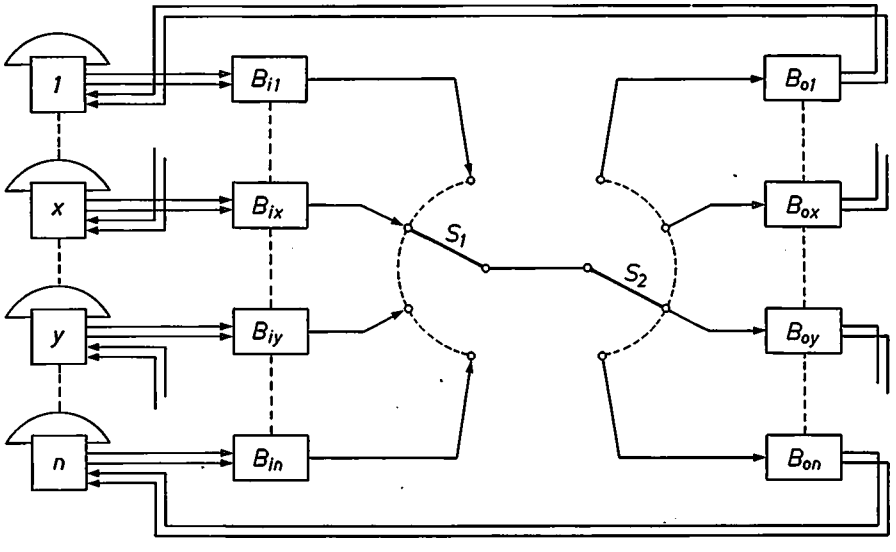


Fig. 4. The principle of time-division multiplex (*TDM*) as applied in type *H* in fig. 3. The signals here are digitized. The bits transmitted by subscriber x are stored in an input buffer B_{ix}. Switch S_1 scans all the input buffers periodically and their contents are transferred via S_2 at an accelerated rate to the appropriate output buffer. In the case of subscriber x, for example, this is output buffer B_{oy} for subscriber y. On leaving the output buffer the signals go to the appropriate subscriber at the original bit rate.

the currents (up to 60 mA) are in fact extremely inconvenient for the introduction of electronics in local exchanges. Electromechanical switches will long continue to be used in these exchanges, since they are well suited to such voltages and currents. On the other hand, electromechanical switches are less suitable for signals with a large bandwidth (video, high-speed data).

The specifications for permissible signal attenuation and the spread in this attenuation are often rather strict. These requirements date from the time when all the switching was electromechanical: then, low values for the magnitude and spread of the attenuation indicated that the switches were in good condition. In view of the attenuation that may occur in the total path from subscriber to subscriber these requirements now seem unrealistic and are even an obstacle to the introduction of electronic circuits into the signal path.

Because of the high sensitivity of the human ear, the requirements relating to crosstalk in telephony are no less exacting, and a crosstalk attenuation of over 70 dB is required for the total connection through the exchange. This is equivalent to a crosstalk attenuation of over 100 dB per crosspoint. For video signals a crosstalk attenuation of 65 dB per crosspoint at 1 MHz is sufficient, but since crosstalk usually increases with frequency by 6 dB/octave, this requirement is by no means less exacting. With digital signals the requirements are much easier.

Certain requirements also have to be met for protection from lightning, and for reliability and facilities for testing. Lightning can produce very high voltages (1000 V) on the lines entering the exchange. These voltages, which are of short duration (1 ms), must not damage the equipment. Protection of integrated circuits is obviously much more difficult than the protection of robust electromechanical and electromagnetic components. The reliability requirements that communication systems have to satisfy are very exacting, particularly in telephony. In some cases they can only be met by duplicating or even triplicating vital elements. Rapid repair and early detection of faulty components necessitate 'testability'. In complicated digital systems this presents great problems.

An SDM network for analog signals with *P-N-P-N* switches

After the above general observations we shall now take a closer look at one particular type of exchange, which has been selected for the experimental videotelephone exchange at Philips Research Laboratories; it is an SDM system for analog signals with electronic switches (type *D* in fig. 3). The switches are *P-N-P-N* semiconductor devices [3].

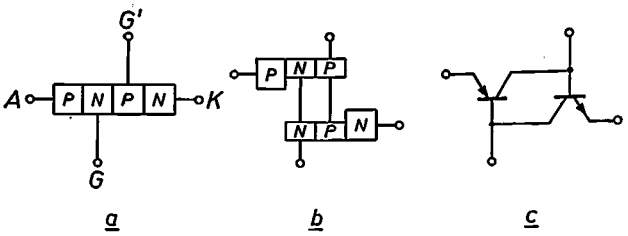


Fig. 5. *a*) A *P-N-P-N* semiconductor element or thyristor. *A* anode. *K* cathode. *G* control electrode. *G'* optional second control electrode. *b*) Hypothetical division into two transistors. *c*) Equivalent circuit diagram. If the element is conducting, it provides its own base currents and thus remains conducting, even if *G* is not connected. If it is not conducting, a short negative pulse at *G* is sufficient to ignite it.

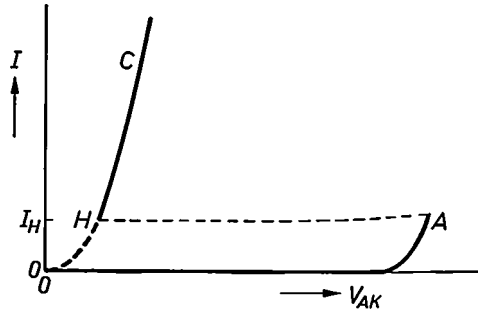


Fig. 6. Current-voltage characteristic of a *P-N-P-N* element. *I* current through the element. V_{AK} voltage between anode and cathode. The characteristic in the absence of a control current is shown by curve *OAHC*. If the voltage V_{AK} reaches point *A* the operating point shifts to the range *HC* and the element becomes conducting. It remains so as long as the current does not fall below the holding current I_H. If the control current is sufficiently large the *P-N-P-N* element assumes the characteristic *OHC*.

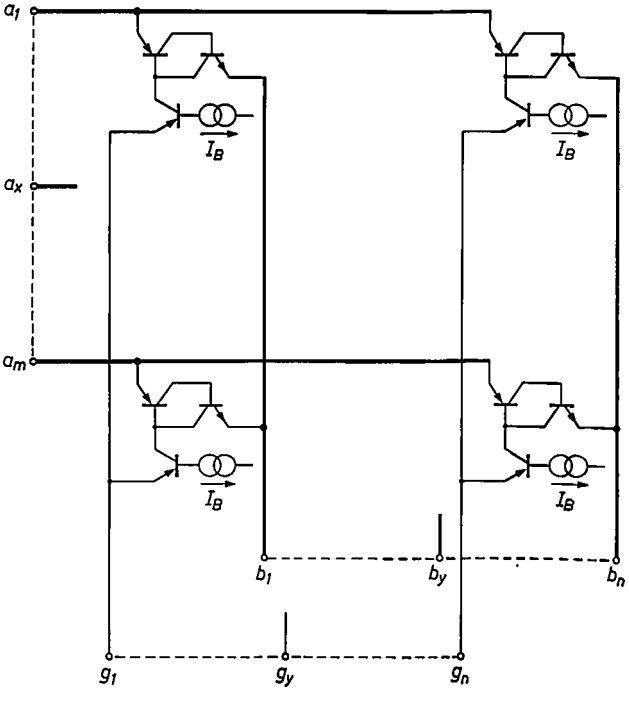


Fig. 7. Matrix of *P-N-P-N* switches. $a_1 \ldots a_m$ incoming lines. $b_1 \ldots b_n$ outgoing lines. $g_1 \ldots g_n$ control lines. Every *P-N-P-N* switch is controlled via a *P-N-P* transistor, so that a control electrode never 'floats'. I_B constant base current of the control transistor.

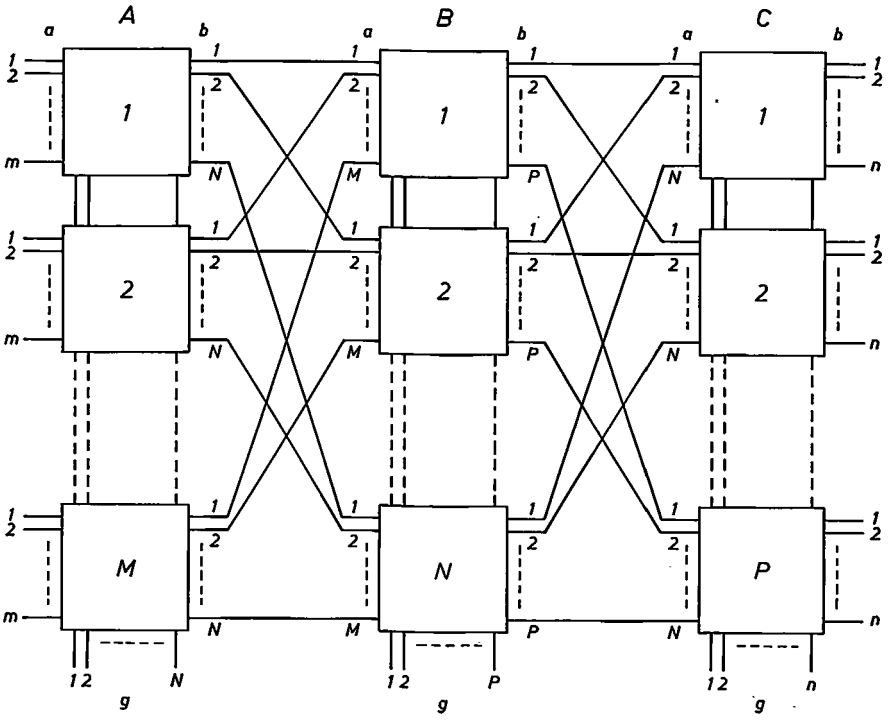


Fig. 8. Switching network for $M \times m$ incoming and $P \times n$ outgoing lines. It consists of three stages A, B and C and is made up of matrices interconnected by 'links'. It is designed to reduce the number of crosspoints needed as compared to a large single matrix with the same number of incoming and outgoing lines.

The P-N-P-N switch

For every switch, whatever the type of exchange, SDM or TDM, there is a memory in which the desired state of the switch is stored. With most types of electronic switches the memory is added separately in the form of a bistable circuit. The *P-N-P-N* switch, however, performs both functions. It is a four-layer semiconductor device, two, three or even four of whose layers can have a connection to the outside (*fig. 5a*). This element is also known as the thyristor (or silicon-controlled rectifier, SCR). It can be considered to be formed from two transistors; see fig. 5*b*,*c*. These themselves provide the base currents required in the conducting state so that this state remains unchanged even without an external control signal. If on the other hand the element is not conducting, a short negative pulse at G is sufficient to make it conducting.

The current-voltage curve of a *P-N-P-N* switch (*fig. 6*) contains two characteristic parts. In the part OA the current is very small; the *P-N-P-N* switch is not conducting. The part HC describes the behaviour of the switch in the conducting state. The change from the conducting to the non-conducting state occurs when the current I is momentarily interrupted. The change from the non-conducting to the conducting state occurs if the voltage V_{AK} is increased to beyond point A (breakdown) but also if a control current is applied momentarily via G or G'. When this happens the current-voltage curve changes from $OAHC$ to OHC (diode characteristic). If the control current is removed, the *P-N-P-N* element remains conducting provided current I remains larger than the 'holding current' I_H. Ignition with a control current is more easily controlled than ignition by breakdown.

A matrix of P-N-P-N switches

The switch described above can be readily used as an element in a matrix (*fig. 7*). An element (x,y) in the matrix is activated by the coincidence of 'marking potentials' on lines a_x and g_y. A negative pulse on the control line g_y has to coincide with a positive pulse on the signal line a_x, which supplies the anode. This signal line is thus used for two functions [4]. If the current through element (x,y) is never allowed to become smaller than the holding current, the marking on a_x and g_y can subsequently be removed. The *P-N-P-N* element is controlled via a transistor, which ensures that the control electrode never 'floats'. If the element is not conducting, the *P-N-P* control transistor is saturated and the control electrode is connected to the g line through a low resistance. When the element is conducting a current βI_B flows to the control electrode (β is the current gain of the transistor). If the element is switched on, so that there are marking potentials on lines a and g, the *P-N-P* control transistor operates in reverse and a current $I_B(\beta_{inv} + 1)$ is drawn from the control electrode. This means that the *P-N-P-N* element must now start to conduct.

[3] E. A. Aagaard, J. W. Coenders and E. C. Dijkmans, *P-N-P-N* interface-to-interface switching for telephony and broad-band communication, IEEE Trans. COM-22, 1250-1263, 1974.

[4] A. Heetman, The switching network in an experimental electronic telephone exchange, Philips Telecomm. Rev. 23, 62-69, 1962.

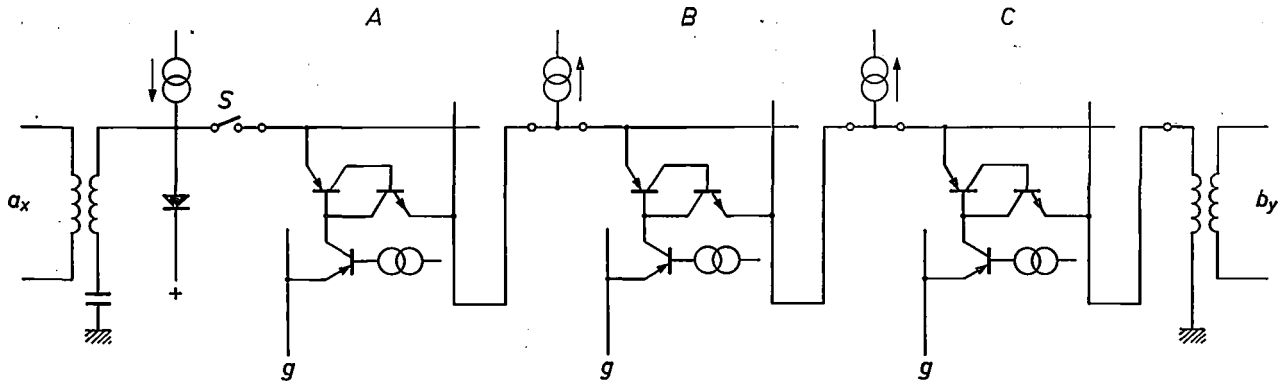


Fig. 9. Connecting path from line a_x to line b_y through a three-stage network with *P-N-P-N* switches; transmission is on two wires and is unbalanced. The three stages are switched simultaneously or sequentially by marking voltages on lines g; in sequential switching the current generators provide the holding current via the circuit stages which are already switched. When switch S opens, the crosspoint switches release.

The control of P-N-P-N switching networks

In small switching networks (up to several tens of inputs and outputs), a single matrix will give economic results. The subscribers can, for example, be connected to the inputs while the outputs are connected to junction circuits each of which links two subscribers' lines (and inserts the ringing signal, ringing tone, etc.). The matrix then connects the calling and the called subscriber to the same junction circuit.

In larger networks more complicated structures have to be used since the number of crosspoints would otherwise become excessive. If, for example, a network for 4000 subscribers (inputs) and 400 outputs consisted of a single matrix, there would be 1.6×10^6 crosspoints. In such an arrangement a connection is not made through a single switching element but through several (*fig. 8*).

If k switching elements are necessary per connection, the network is said to be a k-stage network. In the example given above a 3-stage network would have 327 600 crosspoints. The connecting wires between stages are called 'links'. The processor that controls the network searches for the free links required for each connection. If a network has sufficient links it is non-blocking; as long as one input and one output are free, they can always be interconnected. If the number of links is below a certain minimum, the network is of the blocking type; connections cannot invariably be set up [5].

Fig. 9 illustrates the connecting path through a 3-stage network for unbalanced 2-wire transmission. Only the relevant switching elements are shown. The connection can be made by closing switch S (marking a_x) and by simultaneously presenting a marking voltage at the three desired g wires. In practice the connection is often built up step by step to allow defective crosspoints to be located by means of check measurements. The holding current must then be provided by the two current generators shown between stages *A*, *B* and *C* in fig. 9 until the connection has been made. The connection is broken by opening switch S.

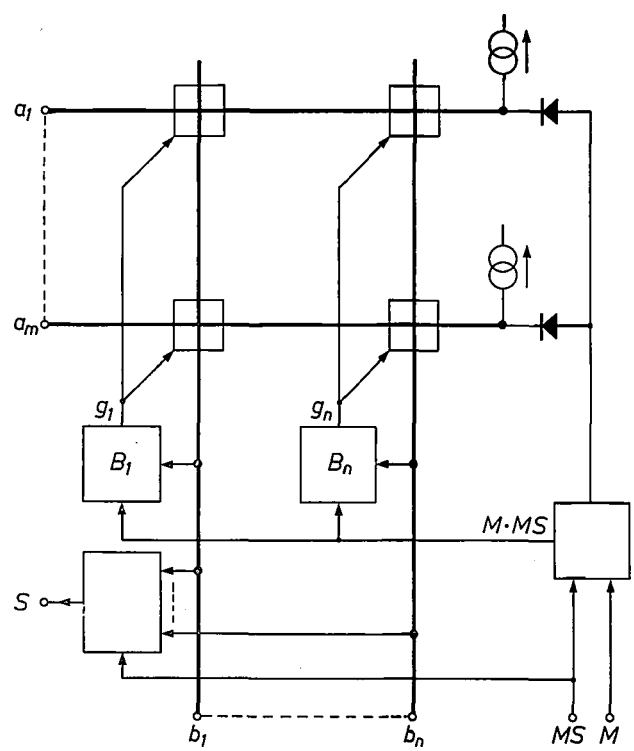


Fig. 10. A matrix in which the g lines are activated via the b lines to reduce the number of contacts. Circuits B_1 to B_n are used for this purpose; they emit a marking pulse when a negative 'search pulse' on the b line coincides with two control signals M and MS. The coincidence of M and MS is indicated by the logic product $M{\cdot}MS$. The control signal MS is also used to send a negative search pulse via the a lines to the b lines of the preceding matrices. S test signal for control purposes.

[5] C. Clos, A study of non-blocking switching networks, Bell Syst. tech. J. 32, 406-424, 1953.

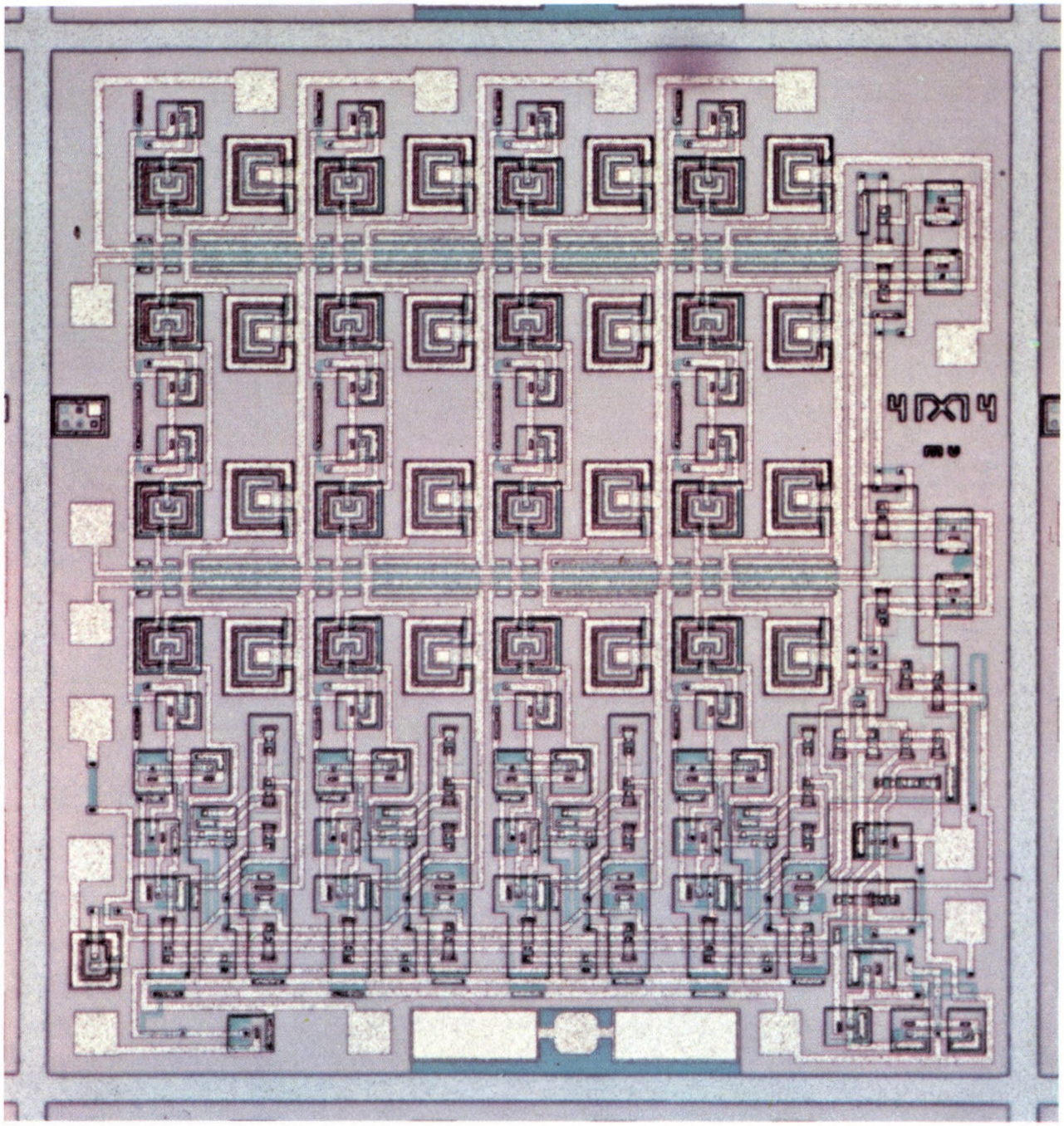


Fig. 11. Matrix of 4×4 *P-N-P-N* elements as an integrated circuit.

Integrated matrices

If a matrix of *P-N-P-N* switches is made as an integrated circuit, it has to have $m + 2n$ contact pins in addition to those required for the power supply (see fig. 7). This is a drawback for large values of n. The problem has been solved by the possibility of selecting the g wire via the associated output (b wire) (*fig. 10*). A characteristic potential sent back from the next stage in the network produces the marking on the g line via the b output and via the associated control circuit B.

An arrangement is therefore incorporated in which the horizontal lines $a_1 \ldots a_m$ of a matrix, which are connected to b lines in the preceding stage, can be brought to this characteristic potential, called the 'search potential'. In fig. 10 this process takes place via input *MS* ('matrix selection'), which distributes a negative pulse across all the horizontals of the matrix that are not in use; the current generator keeps the diode conducting in these lines. If a horizontal is in use it is at a potential higher than earth potential and the diode blocks the negative search pulse.

A search potential is thus transmitted over a number of lines in parallel. Each of these lines leads to another matrix in the preceding stage; see fig. 8. Only in one

of these matrices will the search pulse coincide with a positive marking voltage on an *a* line there, and this will result in the through-connection of a crosspoint. The search potential on the other lines will have no effect.

The circuit of fig. 10 also has an input *M* for applying a marking command. Only when this signal is generated simultaneously with a search command *MS* will an ignition pulse be applied to the *P-N-P-N* elements, and then only to the vertical line on which there is also a negative search potential. Where these conditions are not satisfied the *B* network remains inactive. On the positively marked *a* line the crosspoint starts to conduct; on the other *a* lines the ignition pulse has no effect.

If there is a negative search potential on any *b* line, and a search pulse *MS* is applied to the appropriate matrix, the effect will be to generate a test signal *S*, which will be used by the control system.

We have made integrated matrices of 4×4 *P-N-P-N* devices as shown in fig. 10; see *fig. 11*. This required the use of a specially adapted process. *Fig. 12* shows a network with 16 inputs and 16 outputs in a configuration of the kind illustrated in fig. 8 with $M = N = P = 4$. Matrices larger than 4×4 can be produced easily with available integrated circuits by adding 4×4 modules (*fig. 13*). The *M*, *MS* and *S* control wires of the modules can be interconnected; control of a large matrix is exactly the same as control of a small one.

Matrices with other switching elements

When *P-N-P-N* elements are integrated, the problem of insulating the crosspoints from the substrate arises. An excessive leakage current can readily occur. We have solved this problem by taking an unusual sequence of process steps. If this is considered undesirable and a standard process is preferred, the only remaining possibility is to employ bipolar or field-effect transistors as switches. Since these elements only operate as switches, the memory has to be added. This, however, gives rise to problems relating to the number of pins and the wiring.

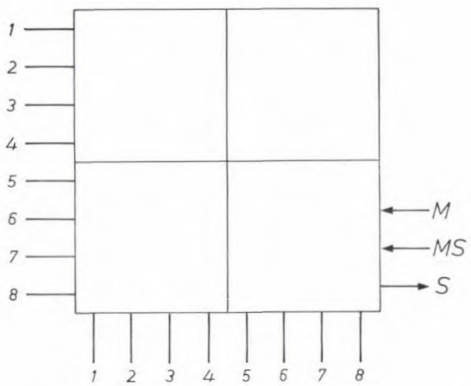


Fig. 13. A number of integrated matrices can be combined to form a larger matrix. This is an 8×8 integrated *P-N-P-N* matrix made up from four 4×4 matrices. The *M* and *MS* inputs and the *S* outputs of the component matrices are interconnected.

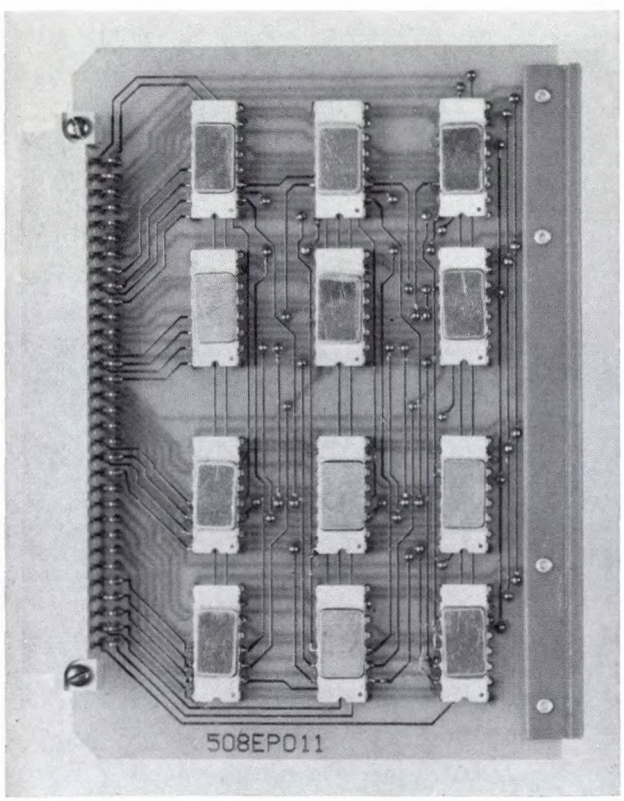


Fig. 12. A three-stage switching network for 16 inputs and 16 outputs with 12 integrated *P-N-P-N* arrays of the kind shown in fig. 11.

Fig. 14 shows a solution to both problems. The information required for selecting a crosspoint is inserted serially in a shift register forming part of the integrated circuit. The desired horizontal and vertical are now marked via two decoders. It is of course also necessary to indicate whether the bistable circuit selected has to be set to the '0' or the '1' state. An extra bit in the shift register or an extra input can be used for this purpose.

Many variations on the control system illustrated in fig. 14 are possible. In most cases an input cannot be connected to more than one output and vice versa. Not every switching element requires a bistable circuit then: it is sufficient to have $\log_2(n+1)$ bistable circuits per row if the number of switches per row is *n*. As in the case of the *P-N-P-N* module, small modules can be joined together to form a larger matrix. The matter is nothowever quite so simple for the control inputs.

Transmission characteristics; crosstalk

Exchanges should introduce very little signal attenuation. With electromechanical switches this is no problem. In fact, signal attenuation is no problem with electronic switches either because amplification is possible. There is spread in the signal transmission, however, because of differences between the electronic components.

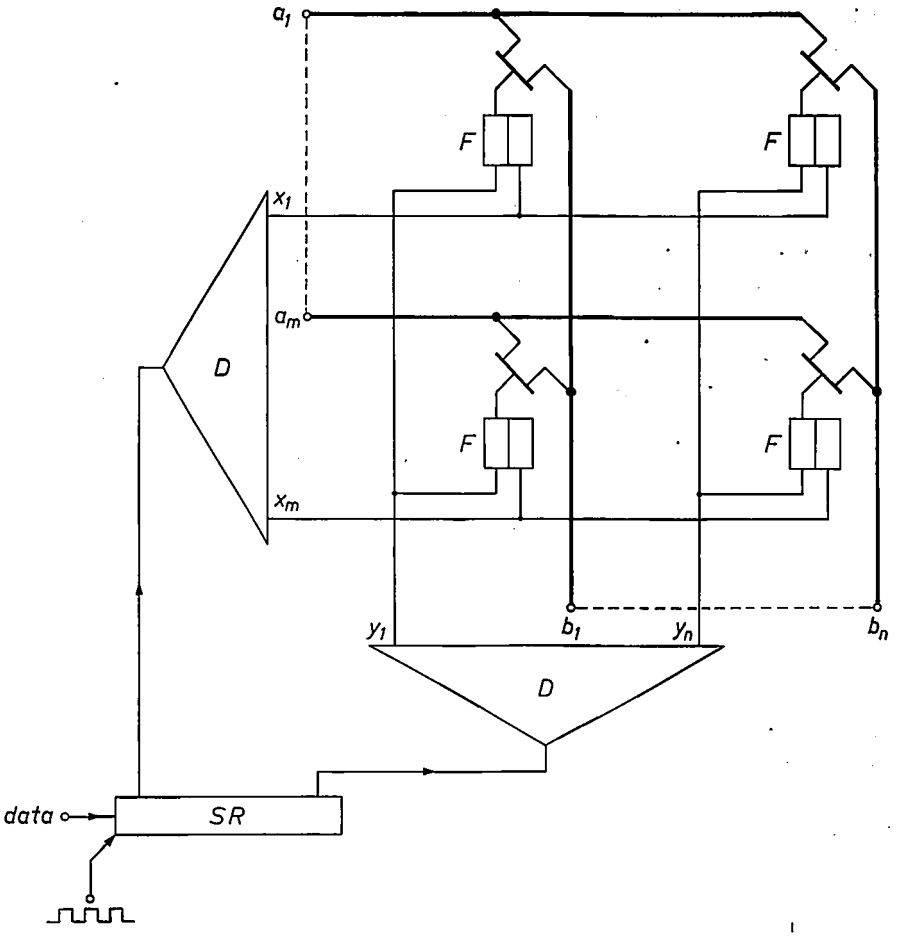


Fig. 14. Matrix with field-effect transistors as switches. The memory function is performed by bistable circuits F that control the transistors. The bistable circuits, in turn, are controlled via decoders D by a shift register SR to which the control information is applied sequentially. This means that with integration the number of external connections can be restricted.

More serious problems are associated with the crosstalk. Because the bandwidth of the signals to be switched in the experimental video-telephone exchange at Philips Research Laboratories was very large, an extensive analysis of the crosstalk was made during the design phase. We shall now present the main points of this analysis.

The circuits in this exchange are unbalanced: they all have a common return path. Inevitably, there are unwanted couplings between the circuits.

We shall now consider a model of two coupled unbalanced circuits (*fig. 15*). The switching network of the exchange is located between P and Q. The transmitting circuits (to the left of P) are represented by a voltage generator of internal impedance Z_i, the receiving circuits (to the right of Q) by a load impedance Z_o. The transmission paths in the exchange are assumed to be short compared with the wavelength of the signals. This means that the couplings can be shown in the form of lumped elements. The couplings appear in various forms: inductive (M), capacitive (C) and resistive (R). We call the crosstalk resulting in the current i_n 'near-end crosstalk', while i_f represents 'far-end crosstalk'. If, for the time being, we neglect R and L in relation

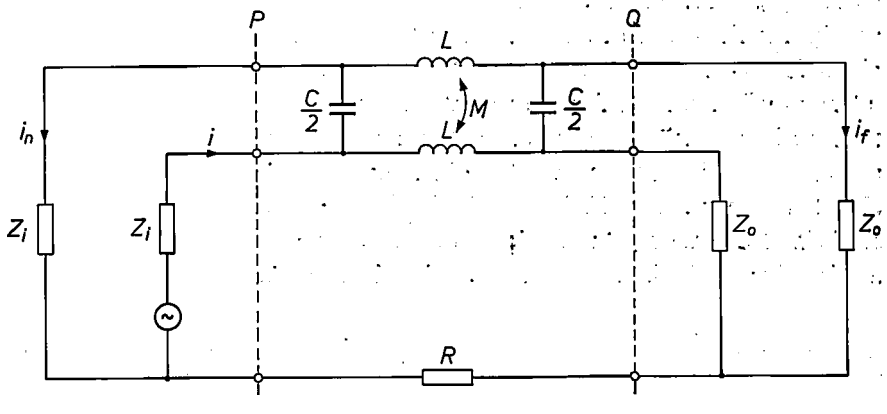


Fig. 15. Model of two (unbalanced) circuits in an exchange, showing their parasitic couplings. The switching network of the exchange is located between P and Q. The couplings are inductive (M), capacitive (C) and resistive (R). The current i in the transmitting circuit sets up unwanted currents in an adjacent circuit both at the input end (near-end crosstalk i_n) and at the output end (far-end crosstalk i_f). The crosstalk can be counteracted by choosing the input impedance Z_i high and the output impedance Z_o low (high-low principle).

to Z_o, it is easily shown that

$$\frac{i_n}{i} = j\omega \frac{CZ_o^2 + M}{Z_i + Z_o}, \quad (1)$$

$$\frac{i_f}{i} = j\omega \frac{CZ_iZ_o - M}{Z_i + Z_o}. \quad (2)$$

In two-wire circuits the following expression is approximately correct:

$$Z_i = Z_o = Z,$$

since the information is conveyed in both directions over the same wires and the sources are identical. Formulae (1) and (2) therefore become:

$$\frac{i_n}{i} = \tfrac{1}{2}j\omega\left(CZ + \frac{M}{Z}\right), \quad (3)$$

$$\frac{i_f}{i} = \tfrac{1}{2}j\omega\left(CZ - \frac{M}{Z}\right). \quad (4)$$

With suitable values of C, M and Z the far-end crosstalk can be low since the capacitive crosstalk and the inductive crosstalk approximately cancel each other out.

In four-wire circuits there is more freedom in the choice of Z_i and Z_o. In addition, it is only the far-end crosstalk that is of significance here because it is the only crosstalk reaching the receiving circuits. We can see from (2) that this can be reduced if the condition

$$Z_iZ_o = M/C \quad (5)$$

is satisfied as closely as possible. It cannot be satisfied exactly, because the values of M and C differ from one circuit to the other. The far-end crosstalk can, however, be kept low by choosing Z_i high and Z_o low, or vice-versa. If (2) is rewritten as

$$\frac{i_f}{i} = j\omega \frac{CZ_o - M/Z_i}{1 + Z_o/Z_i}, \quad (6)$$

it will be clear that the value of the expression will be small if Z_o is low and Z_i high. The adoption of this arrangement has come to be known as the 'high-low principle' [3] [6].

The physical reasons for the advantages offered by the high-low principle are not difficult to explain. The magnetic coupling induces a voltage in the perturbed circuit. If the value of Z_i is high, it ensures that no significant current flows in this circuit. Capacitive coupling is only effective if, as a result of the signal current i, a.c. voltages appear across the coupling capacitances. The lower the value of Z_o, the smaller these a.c. voltages will be. The same reasoning will also make the role of R and L (fig. 15) clear. A resistance R in the common return path causes a voltage iR in the perturbed circuit. The inductance L of the perturbing circuit has the effect of making the voltages at high frequencies large and thus causes an increase in the capacitive crosstalk.

High values of Z_i can in fact be achieved in practice for low frequencies but if the signals have to be supplied from a high impedance to the switching network via several decimetres of wire or printed wiring, Z_i will become increasingly capacitive with increasing frequency:

$$1/Z_i = j\omega C_i + 1/R_i \rightarrow j\omega C_i. \quad (7)$$

The effect of this is to give the crosstalk the frequency response shown schematically in *fig. 16*. Below f_1 the crosstalk is of a resistive nature:

$$\frac{i_f}{i} = \frac{R}{Z_i} \approx \frac{R}{R_i}. \quad (8)$$

Between f_1 and f_2 capacitive coupling predominates:

$$\frac{i_f}{i} = j\omega CZ_o. \quad (9)$$

Above f_2 the crosstalk is given by:

$$\frac{i_f}{i} = \omega^2\left(C_iM - \frac{CL}{2}\right). \quad (10)$$

The term C_iM represents the inductive coupling (Z_i is capacitive); the term $CL/2$ represents the effect of the inductance of the perturbing circuit. In practice f_1 is in the neighbourhood of 200 Hz and f_2 around 5 MHz.

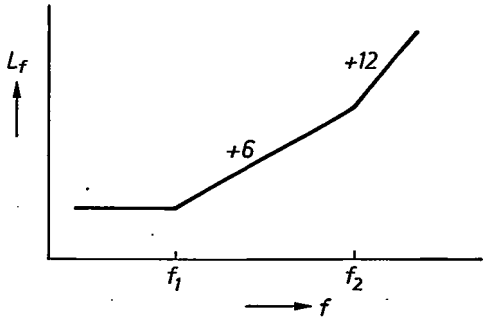


Fig. 16. Representation of the level L_f of the far-end crosstalk as a function of the frequency f. Below f_1 the coupling is mainly resistive, and the crosstalk level is constant; between f_1 and f_2 it is mainly capacitive and the crosstalk increases by 6 dB/octave; above f_2 inductive coupling predominates and the crosstalk increases by 12 dB/octave.

The experimental video-telephone exchange

The high-low principle is employed in the experimental video-telephone exchange at Philips Research Laboratories. This exchange has a three-stage switching network with a capacity of 64 connections. The four-wire transmission for the video telephone means that

every subscriber is connected to the network via a signal input and a signal output. The block diagram in fig. 8 is applicable for $M \times m = P \times n = 64$; it was decided to make $M = P = 4$ and $N = 8$. Because of the small number of links used the network is of the blocking type; as long as the traffic offered is on its present modest scale, this is no disadvantage.

The crosspoints and the control are as shown in fig. 7. A path through the network is indicated in *fig. 17*. Both the input stage and the output stage are formed by transistors in common-base configuration. It was not possible to use the high-low principle in an optimum design because the three *P-N-P-N* crosspoints are in series with the low load impedance; this is why C_1 is the most troublesome of the parasitic coupling capacitances.

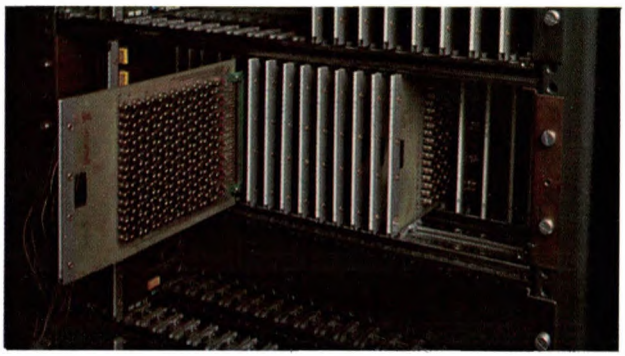


Fig. 18. Detail of the experimental video-telephone exchange at Philips Research Laboratories, Eindhoven. Ten printed-wiring boards carry the entire *P-N-P-N* switching network, which consists of discrete *P-N-P-N* elements. The pulled-out board carries a 16×8 matrix; at each crosspoint there are two small metal 'cans', one for the *P-N-P-N* element and one for the control circuit.

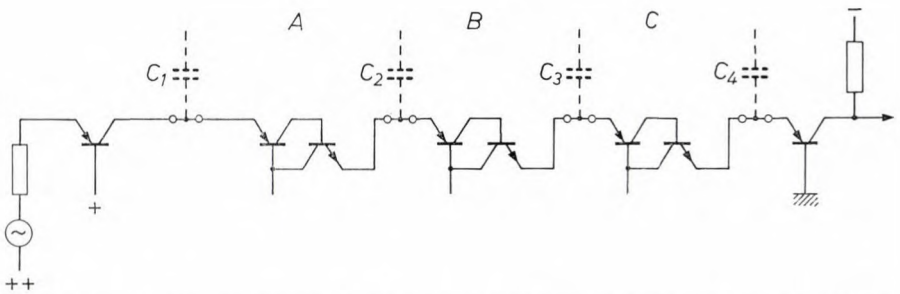


Fig. 17. Connecting path through the three-stage *P-N-P-N* switching network of the experimental video-telephone exchange. The high-low principle is applied in practice by inserting a transistor in a common-base circuit at the input and output. $C_1 \ldots C_4$ parasitic coupling capacitances.

Nevertheless, the video-telephone exchange satisfies the specification for crosstalk. The crosstalk characteristic is shown in fig. 16, although f_2 is high (6 MHz); at 1 MHz the crosstalk attenuation is 55 dB. The frequency response for signal transmission is flat (within 0.5 dB) up to 2 MHz and has a peak at 9 MHz.

The experimental video-telephone exchange is provided with discrete *P-N-P-N* elements. The ten printed-wiring boards on which the entire switching network is accommodated can be seen in *fig. 18*. The board that has been pulled out carries a 16×8 matrix; there are two small metal 'cans' for every crosspoint, one for the *P-N-P-N* element and the other for the control circuit.

Measurements have already been carried out on an experimental network with integrated arrays comprising 4×4 *P-N-P-N* elements and the results are in agreement with those reported above.

Summary. In telephone exchanges of recent design the switches are electromechanical (e.g. reed relays) or electronic. This choice, and that of the form of the signal (analog or digital) and of the switching mode (space division or time division), depends on requirements relating to attenuation, amplitude of signal voltages, protection from lightning, etc. A *P-N-P-N* thyristor can be used as an electronic switch; it has two stable states. We have integrated matrices of 4×4 *P-N-P-N* elements; integrated with these in turn are ancillary circuits that make it possible to ignite the elements via the output lines, so that the number of connecting pins is limited. Crosstalk — a problem particularly in video telephony — can be kept low by taking the signal from a high-impedance source to the switching circuits and giving these a low-impedance load ('high-low' principle). In an experimental video-telephone exchange with *P-N-P-N* switches for 64 subscribers the crosstalk attenuation is 55 dB at 1 MHz.

[6] R. R. Laane, HILO — an improved transmission scheme for semiconductor switching networks, Bell Syst. tech. J. **50**, 1089-1093, 1971.

Recent 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*
Mullard Research Laboratories, Redhill, Surrey, England *M*
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*
MBLE Laboratoire de Recherches, 2 avenue Van Becelaere, 1170 Brussels (Boitsfort), Belgium *B*
Philips Laboratories, 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A. (by contract with the North American Philips Corp.) *N*

Reprints of most of these publications will be available in the near future. Requests for reprints should be addressed to the respective laboratories (see the code letter) or to Philips Research Laboratories, Eindhoven, The Netherlands.

R. N. Alcock, P. England (Royal Aircraft Establishment, Bedford) **& R. P. Vincent:** Performance of the MADGE aircraft guidance system.
Proc. 5th Eur. Microwave Conf., Hamburg 1975, pp. 313-317. *M*

C. Belouet, M. Monnier & R. Crouzier (Université de Paris VI): Strong isotopic effects on the lattice parameters and stability of highly deuterated D-KDP single crystals and related growth problems.
J. Crystal Growth **30**, 151-157, 1975 (No. 2). *L*

P. Blood & R. F. Headon: An automatic system for low temperature electrical measurements on semiconductors.
J. Physics E **8**, 958-963, 1975 (No. 11). *M*

A. Bouwknegt & A. G. van der Kooi (Philips Lighting Division, Eindhoven): Calorimetric measurements of the emission properties of oxide cathodes in a low-pressure gas discharge.
J. Physics D **8**, 952-963, 1975 (No. 8). *L*

R. J. Brewer: The 'barrier mode' behaviour of a junction FET at low drain currents.
Solid-State Electronics **18**, 1013-1017, 1975 (No. 11). *M*

J. J. Brissot & R. Martres: Films minces de polymères organiques pour la détection d'images infrarouges.
Ann. Chimie **10**, 185-194, 1975 (No. 4/5). *L*

E. Bruninx: Characteristics of a NaI(Tl) scintillation chamber.
Int. J. appl. Rad. Isot. **26**, 429-430, 1975 (No. 6/7). *E*

T. M. Bruton: A high temperature oscillating cylinder viscometer.
J. Physics E **8**, 906-908, 1975 (No. 11). *M*

K. H. J. Buschow, M. Brouha & J. B. A. A. Elemans (Reactor Centrum Nederland, Petten): Magnetic order and magnetic history effects in $ThCo_{5x}Ni_{5-5x}$.
Phys. Stat. sol. (a) **30**, 177-186, 1975 (No. 1). *E*

K. H. J. Buschow, W. J. Huiskamp*, H. Th. LeFever*, F. J. van Steenwijk* & R. C. Thiel* (* Kamerlingh Onnes Laboratorium, Leiden): Mössbauer effect and magnetic properties of some Eu-Zn compounds.
J. Physics F **5**, 1625-1636, 1975 (No. 8). *E*

I. M. Coe, J. C. Brice & T. L. Tansley: A statistical barrier model of crystal growth.
J. Crystal Growth **30**, 367-370, 1975 (No. 3). *M*

C. D. Corbey, B. H. Newton & J. G. Summers: Frequency stable M.I.C. S-band and C-band pulsed TRAPATT oscillators.
Proc. 5th Eur. Microwave Conf., Hamburg 1975, pp. 237-240. *M*

J. Cornet: Préparation, caractérisation et propriétés des verres Ge_xTe_{1-x}. Application au stockage optique de l'information.
Ann. Chimie **10**, 239-251, 1975 (No. 4/5). *L*

P. J. Courtois: Decomposability, instabilities, and saturation in multiprogramming systems.
Comm. ACM **18**, 371-377, 1975 (No. 7). *B*

P. J. Courtois: Error analysis in nearly-completely decomposable stochastic systems.
Econometrica **43**, 691-709, 1975 (No. 4). *B*

H. J. van Daal, K. H. J. Buschow, P. B. van Aken & M. H. van Maaren: Possibly mixed valency of uranium in $UNi_{5-x}Cu_x$.
Phys. Rev. Letters **34**, 1457-1460, 1975 (No. 23). *E*

J. J. Daniele: Peltier-induced LPE and composition stabilization of GaAlAs.
Appl. Phys. Letters **27**, 373-375, 1975 (No. 7). *N*

R. Davies & K. L. Fuller: Microwaves today and tomorrow.
Electron (GB) No. 72, pp. 37, 40 & 42, & No. 73, pp. 27, 29 & 30, 1975. *M*

P. Delsarte & J. M. Goethals: Alternating bilinear forms over *GF(q)*.
J. combin. Theory A **19**, 26-50, 1975 (No. 1). *B*

C. Dierieck: Characterization for best nonlinear approximations: a geometrical interpretation.
J. Approx. Theory **14**, 163-187, 1975 (No. 3). *B*

J. W. F. Dorleijn & A. R. Miedema: An investigation of the resistivity anisotropy in nickel alloys.
J. Physics F **5**, 1543-1553, 1975 (No. 8). *E*

J. Drenth, H. M. Swen (both with University of Groningen), **W. Hoogenstraaten & L. A. Æ. Sluyterman:** A mechanism of papain action.
Proc. Kon. Ned. Akad. Wetensch. C **78**, 104-110, 1975 (No. 2). *E*

C. Ducot & F. Serrand (S.A. Philips Industrielle et Commerciale, Paris): Communication et technologie dans la société future.
Revue 2000 No. 31, 2-9, 1975. *L*

J. B. A. A. Elemans*, K. H. J. Buschow, H. W. Zandbergen* & J. P. de Jong* (* Reactor Centrum Nederland, Petten): Magnetic order in the pseudo-binary system $Th(Fe_{1-x}Ni_x)_5$.
Phys. Stat. sol. (a) **29**, 595-600, 1975 (No. 2). *E*

L. J. M. Esser & H. W. Hanneman: Elektronische informatie-overdracht. Revolutie in signaalbewerking.
Natuur en Techniek **43**, 46-59, 1975 (No. 1). *E*

W. G. Essers, G. Jelmorini & G. W. Tichelaar: Soudure par jet de plasma. (*Text to a film produced by* **C. G. Sluyter & P. G. M. Vos.**)
1er Coll. Nat. de Cinématographie Rapide, Arcueil 1974, pp. 83-84; 1975. *E*

L. F. Feiner: Equivalence of propagator and locator approach for the average-*t*-matrix approximation.
Physics Letters **53A**, 349-350, 1975 (No. 5). *E*

E. Fischer, J. Fitzgerald & H. Hörster: Heat and mass transport in gas filled incandescent lamps.
J. Illum. Engng. Soc. **4**, 271-278, 1975 (No. 4). *A*

K. L. Fuller: New radar and navigation systems.
Proc. 5th Eur. Microwave Conf., Hamburg 1975, pp. 13-21. *M*

J. A. Geurst, A. M. J. Spruijt & C. J. Gerritsma: Dynamics of $S = \frac{1}{2}$ disclinations in twisted nematics.
J. Physique **36**, 653-664, 1975 (No. 7/8). *E*

M. Gleria & R. Memming: Charge transfer processes at large band gap semiconductor electrodes: reactions at SiC-electrodes.
J. electroanal. Chem. **65**, 163-175, 1975 (No. 1). *H*

J. J. Goedbloed & M. T. Vlaardingerbroek: Noise and modulation properties of IMPATT-diode amplifiers.
Proc. 5th Eur. Microwave Conf., Hamburg 1975, pp. 685-689. *E*

W. van Haeringen & M. H. van Maaren: Extreem lage temperaturen.
Natuur en Techniek **43**, 109-123, 1975 (No. 2). *E*

P. Harrop, J. Magarshack, A. Rabier & R. Spitalnik (Philips Inbelsa, Sao Paolo): Large signal behaviour of cathode notch stabilized T.E. amplifier devices.
Proc. 5th Eur. Microwave Conf., Hamburg 1975, pp. 699-704. *L*

J. P. Hazan: Application of Titus and Phototitus to image and data processing.
Photonics, editors M. Balkanski & P. Lallemand, publ. Gauthier-Villars, Paris 1975, pp. 227-246. *L*

B. Hill, J.-P. Krumme, G. Much, R. Pepperl, J. Schmidt, K. P. Schmidt, K. Witter & H. Heitmann: Polycube optical memory: a 6.5×10^7 bit read-write and random access optical store.
Appl. Optics **14**, 2607-2613, 1975 (No. 11). *H*

W. J. van den Hoek & E. G. Berns (Philips Lighting Division, Eindhoven): Emission lines in the spectra of halogen incandescent lamps.
Lighting Res. Technol. **7**, 143-146, 1975 (No. 2).

A. Hoekstra, A. Vos (both with Rijksuniversiteit Groningen), **P. B. Braun & J. Hornstra:** The crystal and molecular structures of tetraphenylhydrazine and related compounds at −160 °C, I. The crystal structure and physical properties of tetraphenylhydrazine (TPH).
Acta cryst. **B 31**, 1708-1715, 1975 (No. 6). *E*

B. Hoekstra & V. A. M. Brabers (Eindhoven University of Technology): Magnetocrystalline anisotropy and inversion degree of manganese-ferrous-ferrites.
Solid State Comm. **17**, 249-253, 1975 (No. 3). *E*

W. K. Hofker (Philips Research Labs., Amsterdam Division): Physical aspects of ion implantation.
Acta Electronica **19**, 29-40, 1976 (No. 1).

J. P. M. van Hooff: Konstruktie en toepassing van een vacuümopspanblok.
Glastechn. Meded. **13**, 82-83, 1975 (No. 3). *E*

J.-P. Hurault: Introduction *(to issues on Ion beams in research and technology).*
Acta Electronica **19**, 9-12, 1976 (No. 1). *(In English and in French.)* *E*

B. B. van Iperen: Design and construction of an inexpensive 20 GHz stabilized IMPATT oscillator in microstrip technique.
Proc. 5th Eur. Microwave Conf., Hamburg 1975, pp. 171-175. *E*

M. Jatteau: Techniques of infrared thermography.
Thermography, Proc. 1st Eur. Congr., Amsterdam 1974, pp. 9-24; 1975. *L*

D. Kasperkovitz: A two-phase plasma-coupled static shift register.
IEEE J. **SC-10**, 143-151, 1975 (No. 3). *E*

M. Klinck: Elektronische Regler: Prinzipien und Verhalten.
H_2O **8**, 417-420 & 425-427, 1975 (No. 21). *H*

H. Kuster & P. J. M. Peters: Elektrisch bediende grijpinrichting voor toepassing in vacuüm.
Glastechn. Meded. **13**, 20-22, 1975 (No. 1). *E*

W. Lechner & O. Schob (Philips Lighting Division, Eindhoven): Temperature measurement of filaments above 2500 K applying two-wavelength pyrometry.
Temperature measurement, 1975, Proc. Eur. Conf., Teddington (Inst. Phys. Conf. Ser. No. 26), pp. 297-305. *A*

F. H. de Leeuw & J. M. Robertson: Observation and analysis of magnetic domain wall oscillations in Ga:YIG films.
J. appl. Phys. **46**, 3182-3188, 1975 (No. 7). *E*

J. H. J. Lorteije: A survey of drive methods for gas discharge, light emitting diode, liquid crystal and electrochromic displays.
Nachrichtentechn. Z. **28**, 196-200, 1975 (No. 6). *E*

J. Magarshack: Design and applications of solid state microwave amplifiers.
Proc. 5th Eur. Microwave Conf., Hamburg 1975, pp. 153-167. *L*

D. Mateika, J. Herrnring, R. Rath* & Ch. Rusche* (* Universität Hamburg): Growth and investigation of $\{Gd_{3-x}Ca_x\}[Ga_{2-y-z}Zr_yGd_z](Ga_3)O_{12}$ garnets.
J. Crystal Growth **30**, 311-316, 1975 (No. 3). *H*

F. Meyer & C. J. Loyen: The creep of oil on steel followed by ellipsometry.
Wear 33, 317-323, 1975 (No. 2). *E*

K. H. Nicholas: Applications of ion implantation in semiconductor devices.
Acta Electronica **19**, 95-115, 1976 (No. 2). *(Also in French.)* *M*

W. J. Oosterkamp: X-ray hazards of electronic apparatus.
Electromagnetic compatibility 1975, 1st Symp. & tech. Exhibition, Montreux, pp. 274-276. *E*

G. den Ouden (Philips Welding Development Laboratory, Utrecht), **J. G. Verhagen** (Philips Centre for Technology, Eindhoven) **& G. W. Tichelaar:** Influence of chemical composition on mild steel weld metal notch toughness.
Welding J. **54**, 87s-94s, 1975 (No. 3, Res. Suppl.). *E*

J. Pribetich*, E. Constant*, A. Farrayre & M. Lefebvre* (* Université de Lille): The effect of transferred-electron velocity modulation in high-efficiency GaAs IMPATT diodes.
Proc. 5th Eur. Microwave Conf., Hamburg 1975, pp. 246-250. *L*

H. Rau: Range of homogeneity and defect interaction in high temperature nickel sulfide $Ni_{1-x}S$.
J. Phys. Chem. Solids **36**, 1199-1204, 1975 (No. 11). *A*

U. Rothgordt: Triboelectric charging of thin foils and its application to a new recording principle.
IEEE Conf. Rec. IAS 10th Ann. Meeting, Atlanta 1975, pp. 6-9. *H*

H. Schaumburg (Valvo RHW, Hamburg): Doping uniformity and reproducibility of ion implanted wafers.
Acta Electronica **19**, 141-145, 1976 (No. 2).

H. Schaumburg (Valvo RHW, Hamburg): Application of the ion implantation technology to the production of varicap (varactor) diodes.
Acta Electronica **19**, 147-149, 1976 (No. 2).

J. P. Scott: Electron image projector.
Proc. 6th Int. Conf. on Electron and ion beam science and technology, San Francisco 1974, pp. 123-136; 1975. *M*

M. Sintzoff: Vérification d'assertions pour des fonctions.
Proc. Conf. on Proving and improving programs, Arc-et-Senans, France, 1975, pp. 11-27. *B*

J. L. Sommerdijk & A. Bril: Efficiency of Dy^{3+}-activated phosphors.
J. Electrochem. Soc. **122**, 952-954, 1975 (No. 7). *E*

F. A. Staas, A. P. Severijns & H. C. M. van der Waerden: A dilution refrigerator with superfluid injection.
Physics Letters **53A**, 327-328, 1975 (No. 4). *E*

M. T. Vlaardingerbroek: On the signal dependence of avalanche noise generation.
IEEE Trans. **ED-22**, 309-313, 1975 (No. 6). *E*

L. Vriens & M. Adriaansz: Electronic Raman scattering from Al, Ga, In and Tl atoms.
J. appl. Phys. **46**, 3146-3150, 1975 (No. 7). *E*

J. H. Waszink: Spectroscopic measurements on a high-pressure Na-Xe discharge and comparison with a non-equilibrium calculation.
J. appl. Phys. **46**, 3139-3145, 1975 (No. 7). *E*

W. F. van der Weg & Y. Tamminga (Philips Research Labs., Amsterdam Division): High energy ion scattering as a tool for studying ion implantation.
Acta Electronica **19**, 47-52, 1976 (No. 1).

H. W. Werner: Analysis of implanted layers by means of secondary ion mass spectrometry (SIMS).
Acta Electronica **19**, 53-66, 1976 (No. 1). *E*

G. F. Weston: Materials for ultrahigh vacuum.
Vacuum **25**, 469-484, 1975 (No. 11/12). *M*

J. P. Woerdman & M. F. H. Schuurmans: Spectral narrowing of selective reflection from sodium vapour.
Optics Comm. **14**, 248-251, 1975 (No. 2). *E*

J. M. Woodcock, J. M. Shannon & D. Diguet (RTC La Radiotechnique-Compelec, Caen): Ion implantation into GaAs and relateed compounds.
Acta Electronica **19**, 151-159, 1976 (No. 2). *M*

F. Zernike: Fabrication and measurement of passive components.
Integrated optics (Topics in applied physics, Vol. 7), editor T. Tamir, publ. Springer, Berlin 1975, pp. 201-241. *N*

H. Zijlstra & F. J. A. den Broeder: Structure and coercivity of $SmCo_5$ magnets.
Proc. Int. Conf. on Magnetism, Moscow 1973, Vol. 5, pp. 304-314; 1975. *E*

PHILIPS TECHNICAL REVIEW

VOLUME 36, 1976, No. 11/12

Modulation

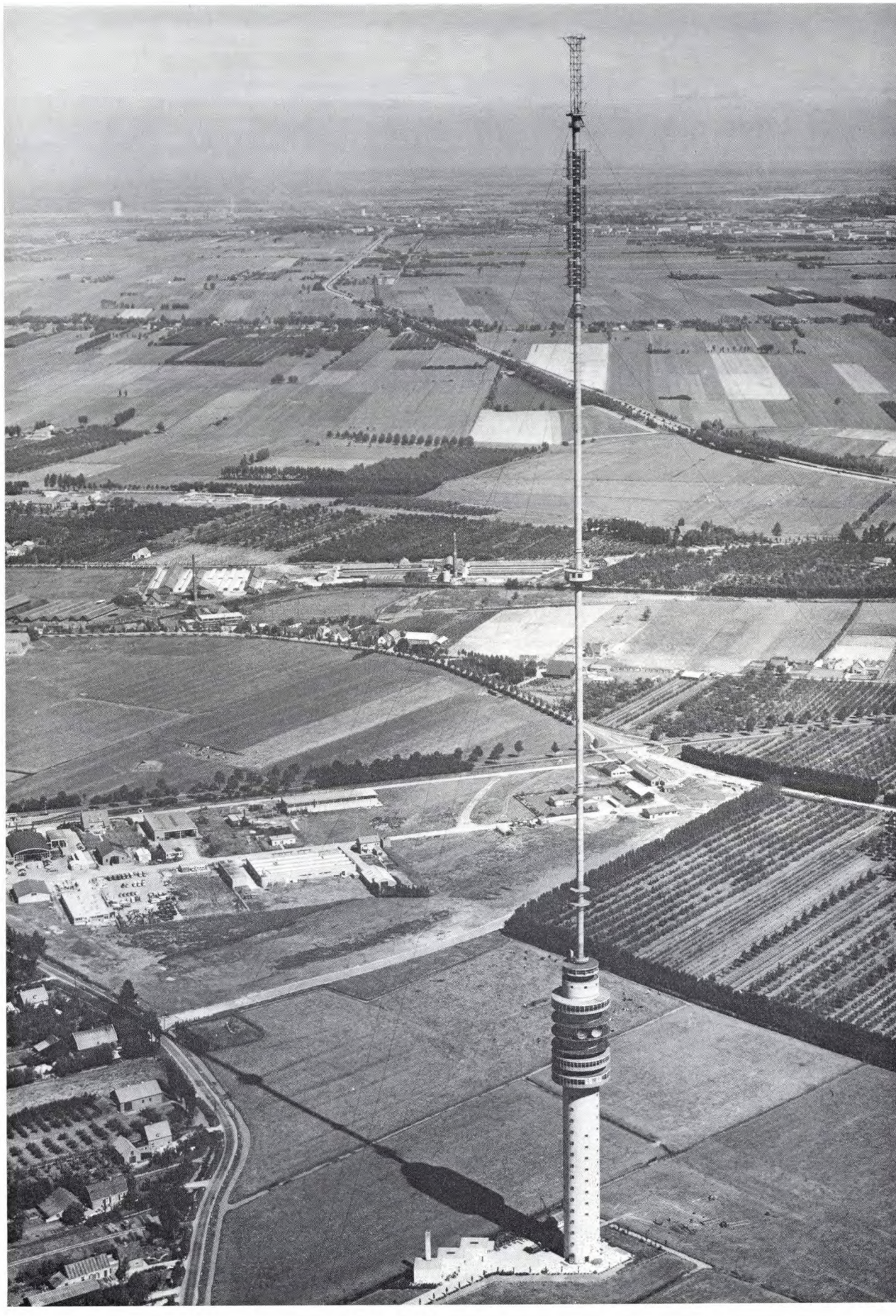
'The difficult problem in wireless telephone is, of course, the modulation of the large amount of energy used for transmission' (R. A. Fessenden, 1907). According to the investigations of R. A. Heising (Proc. IRE 50, 896, 1962) this was the first use of the word 'modulation' as applied to radio. After the first great successes in radiotelegraphy in about 1900, there was much interest in putting the radio waves to work by means of a microphone — a rather more refined instrument than the morse key. Fessenden was undoubtedly thinking here of the technique of amplitude modulation (AM), still in use today. An analysis by J. R. Carson gave more insight into this concept; the existence of sidebands was clearly formulated. Carson's proposal of 'single-sideband modulation with suppressed carrier' was found to be very well suited to applications such as carrier telephony, in which many signals — placed next to one another in the spectrum by modulation ('frequency-division multiplex') — can be transmitted simultaneously over the same cable. The introduction of the thermionic valve, at about the time of the First World War, marked the real start to the development of radiotelephony and carrier telephony. Radiotelephony in turn was the progenitor of radio broadcasting.

The early history of frequency modulation (FM) was rather curious. When the radio spectrum available for broadcasting began to fill up because of the rapid growth in the number of transmitters, many attempts were made to economize in bandwidth by the use of FM, the assumption being that any limitation in the frequency swing would give a corresponding limitation in the bandwidth. This was a misapprehension, however, as Carson showed in 1922; in fact FM requires at least as great a bandwidth as AM. Interest waned, until E. H. Armstrong arrived at the surprising discovery that FM, correctly applied, could give very good suppression of interference (1936). Since that time FM has been one of the most important modulation methods for situations where good signal quality is required and there are no problems with bandwidth.

The interest in the modulation of trains of pulses is related to two possibilities that were put to use in the early days of telegraphy: the combination of pulses from different signals to form a single pulse train ('time-division multiplex') and the regeneration of signals at intermediate stations. Time-division multiplex has certain technical advantages over frequency-division multiplex; signal regeneration prevents the accumulation of noise. The extent to which a given signal can be represented by a train of pulses and a pulse train can be transmitted by a given transmission system was studied in detail by R. V. L. Hartley and H. Nyquist in the twenties. Together with the idea of quantizing the pulses — which is necessary for signal regeneration — this led on to methods such as 'pulse-code modulation' (A. H. Reeves 1939), now widely used in telephony, and more lately taken up for the transmission of pictures (facsimile, television).

Finally, we should mention C. E. Shannon's information theory (1948), which provided the means for studying the extent to which a selected method of modulation can exploit the possibilities of a given transmission channel.

This issue completes our volume for the year that has seen the centenary of the telephone and the 75th anniversary of Guglielmo Marconi's transmission of radio signals across the Atlantic. It seems appropriate to present our readers with a comprehensive survey of the subject of modulation, an essential feature in both telephony and radio, and a field in which Philips have also made important contributions.



Modulation

F. W. de Vrijer

Introduction

In electrical communication systems the information to be transmitted is contained in an electrical signal, called the message signal or information signal. This is the current or voltage, varying with time, which appears at the output of a morse key, a microphone, a television camera or some such other device. When this signal, possibly after amplification, is directly transmitted, by cable for example, we speak of 'baseband transmission'. Examples are telegraphy and telephony by wire and amplifier systems in auditoriums. The 'baseband' is the part of the frequency spectrum that is covered by the information signal.

It was wireless telegraphy that first made use of the fact that high-frequency electromagnetic waves in free space can be received at a great distance from the source. Information can be transmitted on such a 'carrier' by modulating it by the information signal, that is to say by using the signal to vary some characteristic or other of the carrier. It should be noted, incidentally, that carrier transmission is also used today on many cable links.

The 'on-off' method of wireless telegraphy is a primitive form of *amplitude modulation*. With the non-coherent waves from arc transmitters, which were originally used, amplitude modulation was just about the only feasible method. With the coherent, sinusoidal carriers now used, *frequency* and *phase modulation* are used in addition to various forms of amplitude modulation. These are dealt with in part I. When the carrier consists of a train of rectangular pulses, a fourth modulation mode is possible: besides the amplitude, the frequency and the phase of the pulses, their width can also be modulated. Pulse modulation is the subject of part II.

In transmission the frequency of the carrier is very often high compared with the frequencies of the modulating signal; see *fig. 1*. Audio signals contain frequencies up to a maximum of 15 kHz; the lowest frequency band used in broadcast transmission is the familiar long-wave band (150-285 kHz). Video signals for television, on the other hand, contain frequencies of a few MHz, and are transmitted at 'very high' or 'ultrahigh' frequencies (30 MHz and higher). A carrier of particular information may even be transmitted on another carrier of higher frequency, either as an information signal or as part of it. For example, in some television transmitters a carrier at 5.5 MHz is modulated by the audio signal; the resultant modulated signal is then in turn made part of an information signal that also contains video information and is transmitted on a carrier of say 197 MHz.

The transmission covers a region in the frequency spectrum that is situated around the carrier frequency and is at least as wide as the highest frequency of the information signal; the width depends on the method of modulation. Since it is necessary to make the most of the spectrum of frequencies suitable for transmission, the required bandwidth is often one of the main criteria governing the choice of a particular modulation method. Other criteria are concerned with minimizing the effect of signal interferences in the transmission channel. With frequency modulation, for example, it is possible to achieve a higher signal-to-noise ratio than with amplitude modulation, but this does require the use of a wider frequency band.

As the transmission path becomes longer, it is inevitable that noise accumulates in the transmission of continuously varying signals ('analog signals'), until finally the signal is lost in the noise. This can be avoided by converting the analog signal into a train of quantized pulses (a 'digital signal') before transmission. The quantization makes it possible to eliminate the noise from the pulse trains at intermediate repeater stations.

The transmitter mast at Lopik, for both radio and television broadcasting in the Netherlands. The photo was taken in 1962, while UHF transmitting antennas for television band IV were being put into position. The antennas for television band I (VHF, channel 4) were already in position, with the antennas for FM radio beneath them. A number of parabolic antennas for microwave links can be seen on the platforms lower down. With antennas for the national paging system ('Semaphone'), mobile and marine radio ('Mobilophone', 'Mariphone'), the tower now has a height of 420 metres (1380 feet). (Photo KLM Aerocarto.)

Dr F. W. de Vrijer is a Scientific Adviser with Philips Research Laboratories, Eindhoven. Since 1956 he has participated in the work of ITU, CCIR, CCITT, EBU and URSI.

In this conversion process the signal is first 'sampled' in a rapid sequence of time intervals. As a result a pulse train is formed that is modulated in amplitude by the signal. Sampling is therefore a form of pulse modulation, and is dealt with in part II. Next, the pulses are quantized, and as a rule the resultant pulse trains are again transformed into a digital signal, very often a series of ones and zeros (a 'binary signal'). The complete transformation from the analog signal to the digital signal finally transmitted — again a modulated pulse train — is known as the 'coding' of the signal. Quantization and coding are the subject of part III. The whole idea of quantization and coding would of course be pointless if it were not readily possible to transmit the pulse trains produced. The fact that this can be done is by no means self-evident. The problems involved are dealt with in part IV.

It should be added here in passing that the term 'modulation' sometimes denotes *all* transformations of the information signal that are designed to make efficient use of the transmission medium, and thus it also includes the 'coding' we have just referred to. In our use of the word modulation, however, we shall generally keep to the more limited meaning, i.e. 'the variation of characteristics of the carrier or pulse train in accordance with the information signal'.

Modulation techniques are not only used for the transmission of information; they may also be used for the storage and reproduction of information. A typical example is the Philips 'VLP' (video long play) record. Whereas on an ordinary gramophone record the information is stored and is reproduced in the baseband, a 'VLP' record contains a series of small pits which together form a kind of carrier. The information signal is stored as variations in the length and repetition frequency of the pits. Another example is the magnetic recording of video signals.

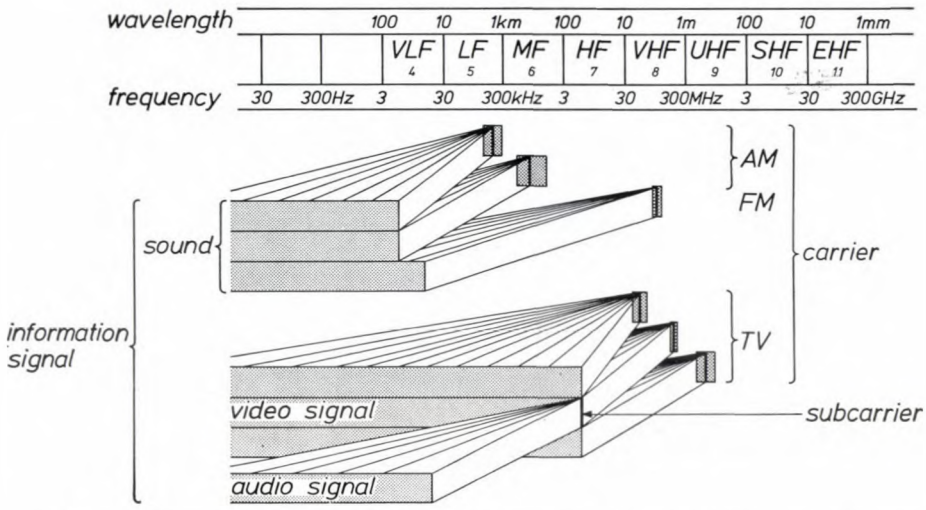


Fig. 1. Division into bands of the part of the electromagnetic spectrum used for telecommunication; frequencies and wavelengths are shown on a logarithmic scale. Some examples of use are given: in AM radio broadcasting sound is usually transmitted with a baseband of 5 kHz, on carriers of 150-285 kHz in the LF (long-wave) band and on carriers of 525-1605 kHz in the MF (medium-wave) band; in FM broadcasting a baseband of 15 kHz is used on carriers of 87-108 MHz in the VHF band. Television signals (baseband several MHz) are transmitted on channels in the VHF or UHF band. An example is also given of a modulated carrier modulating another carrier: in some television transmitters the sound modulates a subcarrier (e.g. at 5.5 MHz), which, together with the video signal, modulates a VHF carrier.

These and other examples discussed in parts I, II and III illustrate the many and various applications of modulation. The pre-eminent field of application, however, is telecommunication. Radio and television are well-nigh unthinkable without modulation, and telephony too would be difficult to imagine nowadays without modulation. In part V we shall review various forms of telecommunication to illustrate the use of modulation in this specific field.

* *

*

I. Modulation of a sinusoidal carrier

Conventional amplitude modulation

The oldest and also the most obvious form of modulation is what we shall refer to as conventional amplitude modulation. In this form the pattern of the information signal is directly 'impressed' on the carrier (*fig. 2a*). In other words, the modulated carrier has an 'envelope' in the shape of the information signal. If we suppose that the carrier transmitter consists of an *LC* circuit in which an electrical oscillation is generated by a transmitter tube, then the modulation can be effected, for instance, by varying the anode voltage of the tube in accordance with the information signal. Recovering the information signal from the modulated carrier at the receiver end — the process of detection or demodulation — is a particularly simple matter in the case of conventional amplitude modulation. A rectifier (diode), followed by an *RC* filter that passes the frequencies of the information signal but not those of the carrier, delivers the desired result (fig. 2*b*). This is known as envelope detection or peak detection.

Let the unmodulated carrier be $A \cos(\omega_c t + \phi)$ and the information signal be $p(t)$; then the modulated signal $S(t)$ has the form:

$$S(t) = \{C + p(t)\} \cos(\omega_c t + \phi). \qquad (1)$$

The constant C has an essential function: it must be large enough for $C + p(t)$ to be positive at all times, for only then will peak detection yield the original signal; see *fig. 3*. If the average value of $p(t)$ is zero, the value of C in the above method of modulation is equal to the average anode voltage. The requirement that $C + p$ shall always be positive may also be expressed by stipulating that the modulation depth m shall not be greater than 1. The modulation depth is defined by the relation (see *fig. 4*):

$$m = \frac{(C+p)_{max} - (C+p)_{min}}{(C+p)_{max} + (C+p)_{min}} = \frac{p_{max} - p_{min}}{2C + p_{max} + p_{min}}. \qquad (2)$$

Often p_{min} is equal to $-p_{max}$, in which case m is equal to p_{max}/C.

To determine the spectrum of $S(t)$, let us first assume that $p(t)$ is sinusoidal:

$$p(t) = a \cos \omega_p t.$$

For $S(t)$ we then find:

$$S(t) = C \cos(\omega_c t + \phi) + \tfrac{1}{2}a \cos\{(\omega_c + \omega_p)t + \phi\} + \tfrac{1}{2}a \cos\{(\omega_c - \omega_p)t + \phi\}. \qquad (3)$$

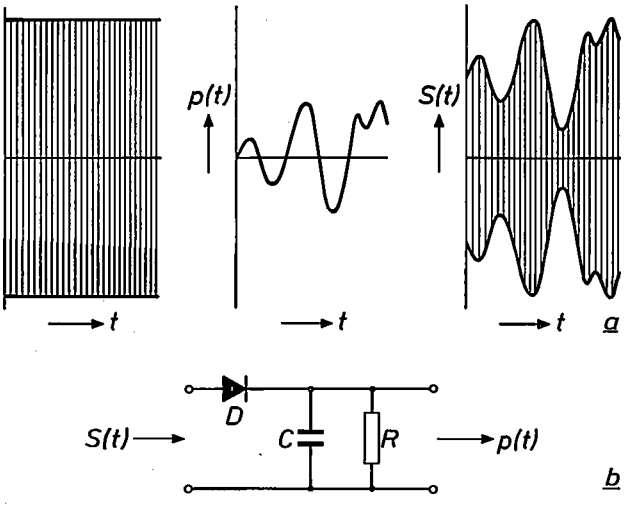


Fig. 2. *a*) Unmodulated carrier (left), information signal $p(t)$ and modulated carrier $S(t)$ in conventional amplitude modulation. The carrier is schematically represented by vertical lines, with the envelope of the peaks drawn as a solid line. *b*) Diagram of peak detector in its simplest form: a diode *D* followed by a lowpass filter *RC*.

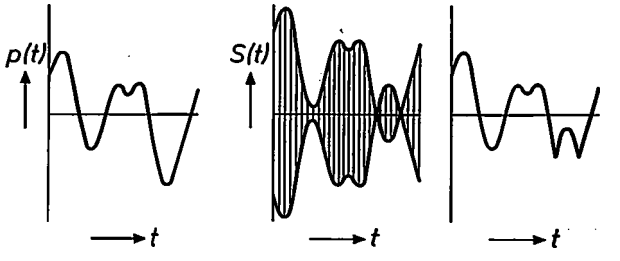


Fig. 3. Information signal $p(t)$, modulated signal $S(t)$ and detected signal (right) for the case where C in eq. (1) is too small, so that $C + p(t)$ now and then changes sign. In peak detection this leads to 'folding over' of the parts of $p(t)$ where $C + p(t)$ is negative.

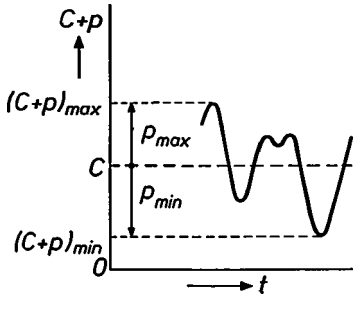


Fig. 4. Illustrating the definition (2) of modulation depth. A modulation depth of 100% ($m = 1$) is only just permissible. Then $(C + p)_{min} = 0$, so that $C + p$ only just fails to change sign.

In this simple case the spectrum thus consists of three lines, at the frequencies $f_c - f_p$, f_c and $f_c + f_p$, with $f_c = \omega_c/2\pi$, $f_p = \omega_p/2\pi$. The vector diagram in *fig. 5*, which rotates with the carrier vector, shows the three components of the signal $S(t)$ of eq. (3): a stationary

vector of amplitude C, and two vectors of amplitude $\frac{1}{2}a$ that rotate in opposite directions at an angular velocity ω_p. The total vector sum does not rotate, but varies only in amplitude, in accordance with eq. (1).

When expressions like those in eq. (3), consisting of a number of cosine terms, are represented in a vector diagram, each term is regarded as the real part of an exponential given by $\cos\psi = \mathrm{Re}\exp(j\psi)$, and the appropriate exponential functions are divided by $\exp(j\omega_c t)$ and plotted as vectors in the complex plane. The effect of this division is to make the carrier vector stationary. Movement of a vector indicates a frequency difference from the carrier.

A periodic but non-sinusoidal information signal can be expanded in a Fourier series, and each component in $S(t)$ then gives rise to a sum frequency and a difference frequency relative to f_c. The total spectrum thus has the character of *fig. 6*: a central line at f_c with two symmetrically located sidebands. If, for example, we use a carrier of 35 000 Hz to transmit telephony speech signals (300-3400 Hz), then the lower sideband occupies the band from 31 600 to 34 700 Hz, and the upper sideband the band from 35 300 to 38 400 Hz.

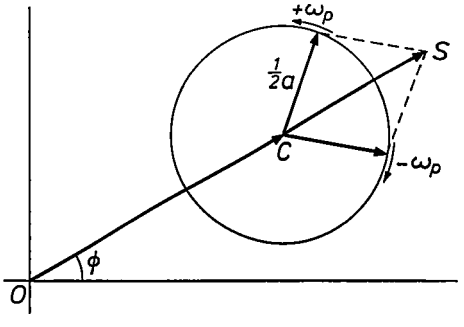


Fig. 5. Vector diagram for conventional AM with a sinusoidal information signal. The vectors in the complex plane represent the exponentials whose real parts are the cosine terms in (3), but only after division by $\exp(j\omega_c t)$, so that the carrier vector OC is stationary. The vectors of the two 'sideband signals' have the same magnitude and rotate in opposite directions at an angular velocity ω_p. The total vector sum OS varies in magnitude but not in direction.

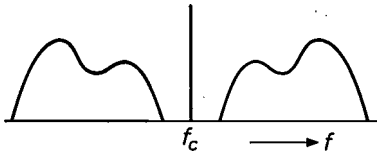


Fig. 6. Spectrum of a modulated carrier in conventional AM, consisting of the carrier-wave component and two sidebands.

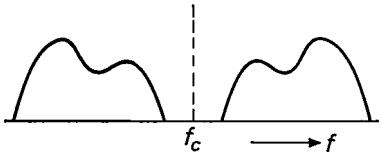


Fig. 7. Spectrum in double-sideband modulation. The carrier-wave component is suppressed, so that the spectrum only consists of two sidebands.

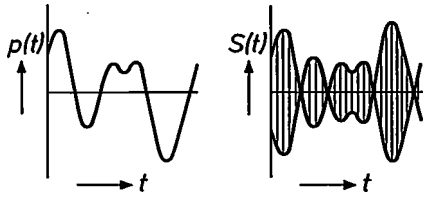


Fig. 8. Information signal $p(t)$ and modulated signal $S(t)$ in DSB modulation.

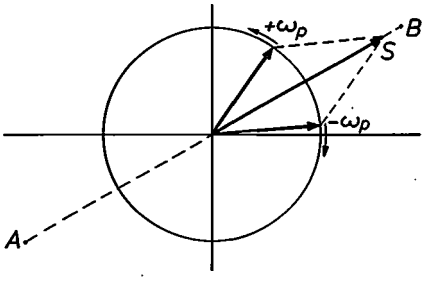


Fig. 9. Vector diagram for DSB modulation with a sinusoidal information signal. The two sidebands rotate in opposite directions at an angular velocity ω_p. The point S corresponding to the vector sum oscillates between A and B.

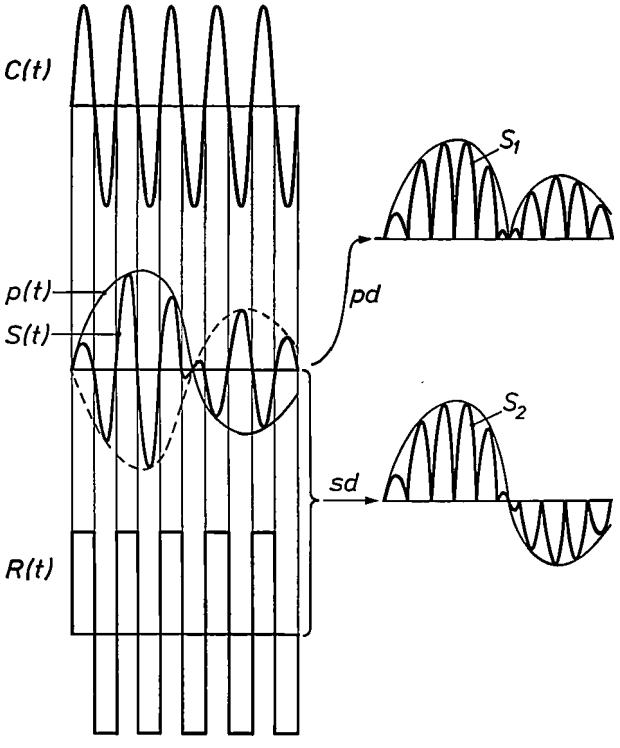


Fig. 10. Peak detection (*pd*) and synchronous detection (*sd*) of the modulated signal (4), where the information signal changes sign. $C(t)$ carrier, $p(t)$ information signal, $S(t)$ modulated signal, $R(t)$ reference signal. The (full-wave) rectification that takes place in peak detection, followed by filtering, gives the signal S_1 which reproduces the positive parts of $p(t)$ correctly but reverses the negative parts. In synchronous detection, $S(t)$ is multiplied by a signal $R(t)$ with the same frequency and phase as $C(t)$. In the figure $R(t)$ is a square-wave signal; after filtering, a signal $S_2(t)$ is then obtained which also reproduces the negative parts of $p(t)$ correctly.

Double-sideband modulation with suppressed carrier (DSB)

The component $C\cos(\omega_c t + \phi)$ in the modulated signal $S(t)$ is in a sense a nuisance and redundant; it takes up a high proportion of the power, but carries no information. In conventional amplitude modulation, however, it cannot be left out, because, as we have seen, the constant C must be sufficiently large to give faithful peak detection.

We shall now discuss a method in which the modulated signal has the form

$$S(t) = p(t)\cos(\omega_c t + \phi). \quad (4)$$

The inconvenient unwanted component is thus eliminated (*fig. 7*), provided $p(t)$ is on average equal to zero. In what follows we shall assume that this is invariably the case, an assumption that is always permissible for audio signals. This method is called 'double-sideband suppressed-carrier modulation' (DSB-SC). For brevity we shall refer here simply to DSB modulation, although it is also amplitude modulation, while on the other hand the same two sidebands occur as in conventional AM. A DSB-modulated signal $S(t)$ is shown in *fig. 8*, and *fig. 9* gives the vector diagram for the case of a sinusoidal information signal.

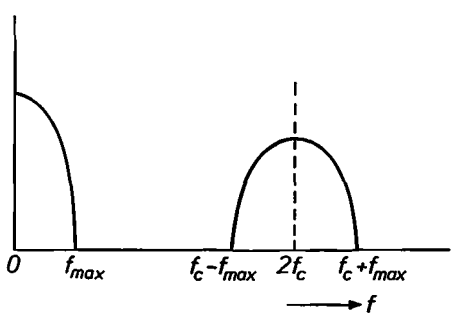


Fig. 11. Spectrum of the signal (5), i.e. the synchronous detection signal after the multiplication of S by R but before the lowpass filter. The low-frequency part is identical with the spectrum of $p(t)$ itself. The spectrum of $p(t)$ can be completely recovered by filtering only if the maximum frequency f_{max} of $p(t)$ is lower than the carrier frequency f_c.

Fig. 10 shows in more detail why the signal of eq. (4) is not suitable for peak detection. Every time $p(t)$ goes through zero the phase of $S(t)$ rotates by 180°, and peak detection is unable to sense this phase reversal. It can however be sensed with *synchronous detection*: $S(t)$ is multiplied by a *synchronous reference signal* $R(t)$, which has the same frequency as the carrier [1]. Fig. 10 shows that with this method, after eliminating the r.f. components with a lowpass filter, the negative parts of $p(t)$ are also faithfully reproduced.

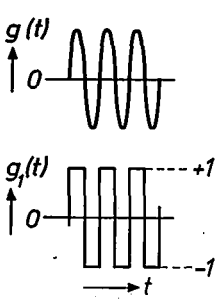
Taking a sinusoidal signal of phase ψ for $R(t)$, e.g. $2\cos(\omega_c t + \psi)$, the multiplication gives:

$$R(t)S(t) = 2\cos(\omega_c t + \psi)\,p(t)\cos(\omega_c t + \phi) = \\ = p(t)\cos(2\omega_c t + \phi + \psi) + p(t)\cos(\phi - \psi). \quad (5)$$

After elimination of the first high-frequency term, $p(t)$ is indeed obtained, but multiplied by $\cos(\phi - \psi)$. The reference signal must therefore be in phase with the carrier ($\psi = \phi$, $\cos(\phi - \psi) = 1$). Phase differences lead to smaller output signals; in the worst case — a phase difference of 90° — the output signal is in fact zero. *Small* phase differences, however, are not very troublesome because the cosine only begins to differ appreciably from 1 at larger values of $\phi - \psi$. *Fig. 11* gives the spectrum of (5), the signal after multiplication but before the first term has been filtered out. It can be seen that the whole process can only be completely successful if the information signal $p(t)$ contains no frequencies higher than the carrier frequency f_c.

In both modulation and detection a particular signal ($p(t)$ or $S(t)$) has to be multiplied by a sinusoidal signal. *Fig. 12* shows how this multiplication is often carried out. Let $f(t)$ be the first signal and $g(t)$ the sinusoidal signal: $g(t) = \sin \omega_c t$. The signal $g(t)$ is applied at high amplitude. In one half-cycle of $g(t)$ both D_1 and D_2 conduct while D_3 and D_4 are biased off; the signal $f(t)$ then arrives directly at the output. In the other half-cycle D_3 and D_4 conduct while D_1 and D_2 are off, and

[1] Instead of 'synchronous' the term 'coherent' is sometimes used.



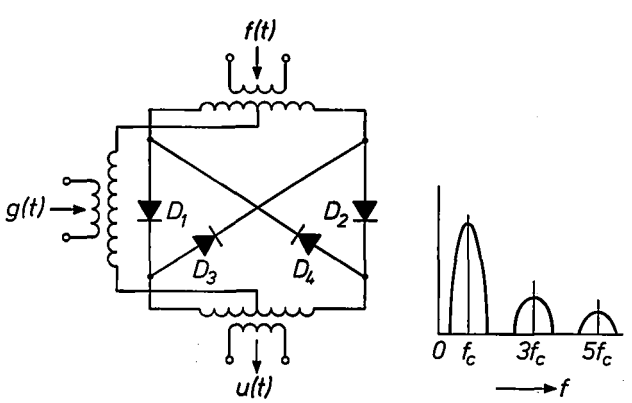


Fig. 12. Modulator for multiplying an arbitrary signal $f(t)$ by a sinusoidal signal $g(t)$. D_1, D_2, D_3, D_4 are diodes, $u(t)$ is the output signal. If the amplitude of $g(t)$ is made sufficiently large, then $u(t)$ is equal to $g_1(t)f(t)$, where $g_1(t)$ is the square wave with the phases of $g(t)$. *Right:* the spectrum of $u(t)$ for the case where $f(t)$ is a low-frequency signal. The first band represents the desired product.

$f(t)$ arrives at the output with the opposite sign. The signal $u(t)$ at the output is thus $f(t)g_1(t)$, where $g_1(t)$ is the square-wave function in fig. 12. Expansion of $g_1(t)$ as a Fourier series yields for $u(t)$:

$$u(t) = f(t)g_1(t) = \frac{4}{\pi} f(t)\{\sin \omega_c t + \tfrac{1}{3}\sin 3\,\omega_c t + \ldots\}.$$

The first term gives the required result, the others can be eliminated by a lowpass filter. Here again, the intended result can only be obtained provided $f(t)$ contains no frequencies higher than f_c (see the spectrum of $u(t)$ in fig. 12).

DSB modulation compared with conventional AM

The quantitative advantage of the suppressed carrier in DSB modulation is that the peak power necessary for good transmission of a given information signal is four times lower with DSB modulation than with conventional modulation. This follows immediately from the fact that the peak amplitude of the modulated signal in conventional amplitude modulation must be at least twice as great as in DSB; see *fig. 13*. In practice it is especially important that this difference of 6 dB in peak power [2] also occurs if a given signal-to-noise ratio after detection is to be achieved. This follows from what has been said above, provided that the effect of noise or other interfering signals incident at the detector is identical for synchronous and peak detection. We shall now show that this is in fact the case.

We assume that the sideband amplitudes resulting from conventional AM or from DSB modulation with a sinusoidal signal are $\frac{1}{2}a$ at the detector. We compare this with the situation where the detector is affected by an interfering signal $b \cos \omega_s t$, at a frequency f_s in the frequency band of the receiver. *Fig. 14* gives the two vector diagrams for conventional modulation; in the second diagram the vector b rotates at an angular velocity $\omega_s - \omega_c$. Peak detection yields the variation in length of the resultant vector OS. This is $a \cos \omega_p t$ in the 'signal' case, and supposing that b is much smaller than the amplitude of the carrier, $b \cos(\omega_s - \omega_c)t$ in the 'noise' case. The signal-to-noise ratio after detection is thus a/b.

In DSB modulation with synchronous detection, multiplication by the reference signal $2 \cos \omega_c t$ yields $2a \cos \omega_p t \cos^2 \omega_c t$ and $2b \cos \omega_s t \cos \omega_c t$; the result of the detection is the low-frequency component

$$a \cos \omega_p t \quad \text{and} \quad b \cos(\omega_s - \omega_c)t,$$

the same result as with peak detection.

If the interference has the character of white noise, the same reasoning applies to each Fourier component of the noise. The high-frequency noise is converted into low-frequency noise in the same way in both peak and synchronous detection.

If the noise is in fact not much weaker than the signal, the above reasoning remains valid for synchronous detection. For peak detection, however, complications arise that result in a relatively lower signal-to-noise ratio.

High-frequency white noise in a frequency band around f_c can be described as two uncorrelated low-frequency noise signals, $R_1(t)$ and $R_2(t)$ of half the bandwidth, which DSB modulate two (suppressed) carriers at frequency f_c 'in quadrature' (quadrature modulation, see p. 317):

$$N = R_1(t) \cos \omega_c t + R_2(t) \sin \omega_c t.$$

In the case of synchronous detection with reference signal $\cos \omega_c t$ the detected signal is proportional to $R_1(t)$ but insensitive to $R_2(t)$, even when $R_1(t)$ and $R_2(t)$ are large.

Peak detection in the presence of a carrier of amplitude A likewise only yields $R_1(t)$ and not $R_2(t)$ if $R_1(t)$ and $R_2(t)$ are much smaller than A. If this is not the case, however, $R_2(t)$ starts to make a considerable indirect contribution. This follows from the fact that on detection — a nonlinear operation — noise components with a frequency difference Δ constitute a low-frequency mixing product of frequency Δ.

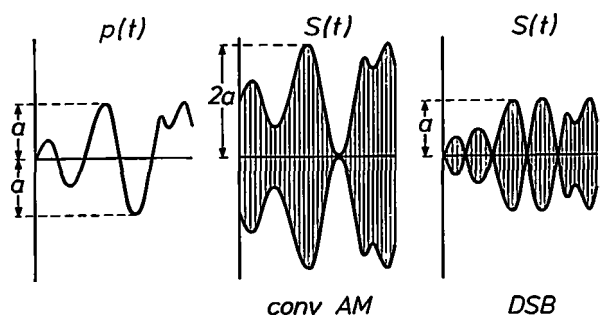


Fig. 13. Peak power in conventional AM and DSB modulation. In conventional AM with 100% modulation depth the peak amplitude of the modulated signal $S(t)$ is twice the maximum a of the information signal $p(t)$, provided that the maximum and the minimum of $p(t)$ are identical in absolute value. In DSB modulation the peak amplitude is only a.

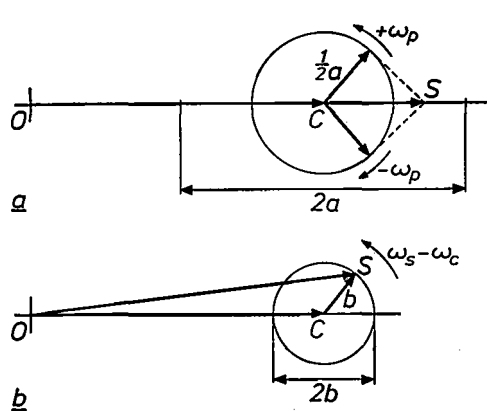


Fig. 14. Vector diagrams for conventional AM. OC carrier. *a*) Sinusoidal information signal with amplitude a ($a < OC$); *b*) no information signal, but an interfering signal $b \cos \omega_s t$ at the detector ($b \ll OC$). The vector b rotates at an angular velocity $\omega_s - \omega_c$. Peak detection yields the variations in the length of the vector sum OS. The amplitudes of the variation are a and b respectively.

Although DSB modulation has a 6-dB advantage in the peak power required, it has the disadvantage that the detection is much more complicated. The receiver must contain a reference oscillator that is synchronous with the unmodulated carrier. The phase information is usually sent from the transmitter to the receiver in the form of a pilot signal, independently of the modulated signal. This is a signal of low amplitude with a fixed frequency and phase relationship to the suppressed carrier. In the simplest case the frequency and phase are identical to those of the original carrier, but this is not necessarily so. The phase may for example be shifted by 90°. There may also be a difference between the pilot and carrier frequencies: however, this should be in the ratio of two (small) integers.

Briefly, the practical difference between DSB modulation and conventional AM thus amounts to the fact that DSB modulation permits the use of a smaller transmitter, while conventional AM permits much simpler detection. In each application it is necessary to decide which carries the most weight. In broadcast systems, for example, there is a single transmitter and a very large number of receivers. For this reason conventional AM is still used for (AM) broadcasting: the need for a larger transmitter is of negligible importance compared with the advantage of simple peak detection.

In two-way radio communication between fixed stations, between mobile units such as ships, aircraft, taxis and police cars, or between such units and a fixed station, the situation is completely different. Here each communicator has a transmitter, and therefore the size of the transmitter is very important and the advantage of carrier suppression predominates. Nevertheless, DSB modulation is not much used today, since another method is available that offers much the same advantages but takes up less bandwidth. This method is single-sideband modulation.

Single-sideband modulation (SSB)

The two sidebands of the DSB signal (fig. 7) each contain the complete information, and to economize on bandwidth it is obvious that one of them could be suppressed completely, e.g. by means of a filter. This is the idea behind single-sideband modulation.

The suppression of a sideband has important consequences for the detection. With a sinusoidal information signal

$$p(t) = a \cos \omega_p t,$$

the modulated signal after suppression of the upper sideband is:

$$S(t) = \tfrac{1}{2}a \cos \{(\omega_c - \omega_p)t + \phi\}.$$

Whereas in fig. 5 and fig. 9 the vector $S(t)$ changed in magnitude but not direction, we now have the opposite situation (*fig. 15*). The result of synchronous detection with $R(t) = 4 \cos (\omega_c t + \psi)$ as reference signal is $a \cos (\omega_p t + \phi + \psi)$. A phase error in the reference signal thus produces a phase shift and not, as in DSB modulation, a change in the amplitude of the detected signal. If the information signal is not sinusoidal, then the effect of a phase error is that all the Fourier components receive a phase shift but maintain their amplitude. In DSB modulation a phase error had the effect of reducing the whole signal.

For audio signals such phase shifts, which imply relative phase shifts between the components, are permissible. This is because the human ear behaves rather like a detector that resolves the acoustic signals into Fourier components and records only the amplitude. Speech remains intelligible even with continuous slow changes in phase differences, which means that a small difference in frequency between carrier and reference signals is admissible. The pilot signal can then be omitted and a stable local oscillator at the receiver can provide the reference signal. The tolerance in the reference frequency for good speech intelligibility is many tens of hertz. The ITU (International Telecommunication Union) standard for good telephony is a maximum deviation of 20 Hz. Such frequency shifts are not permissible in the transmission of music.

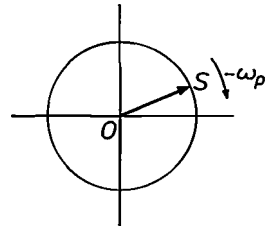
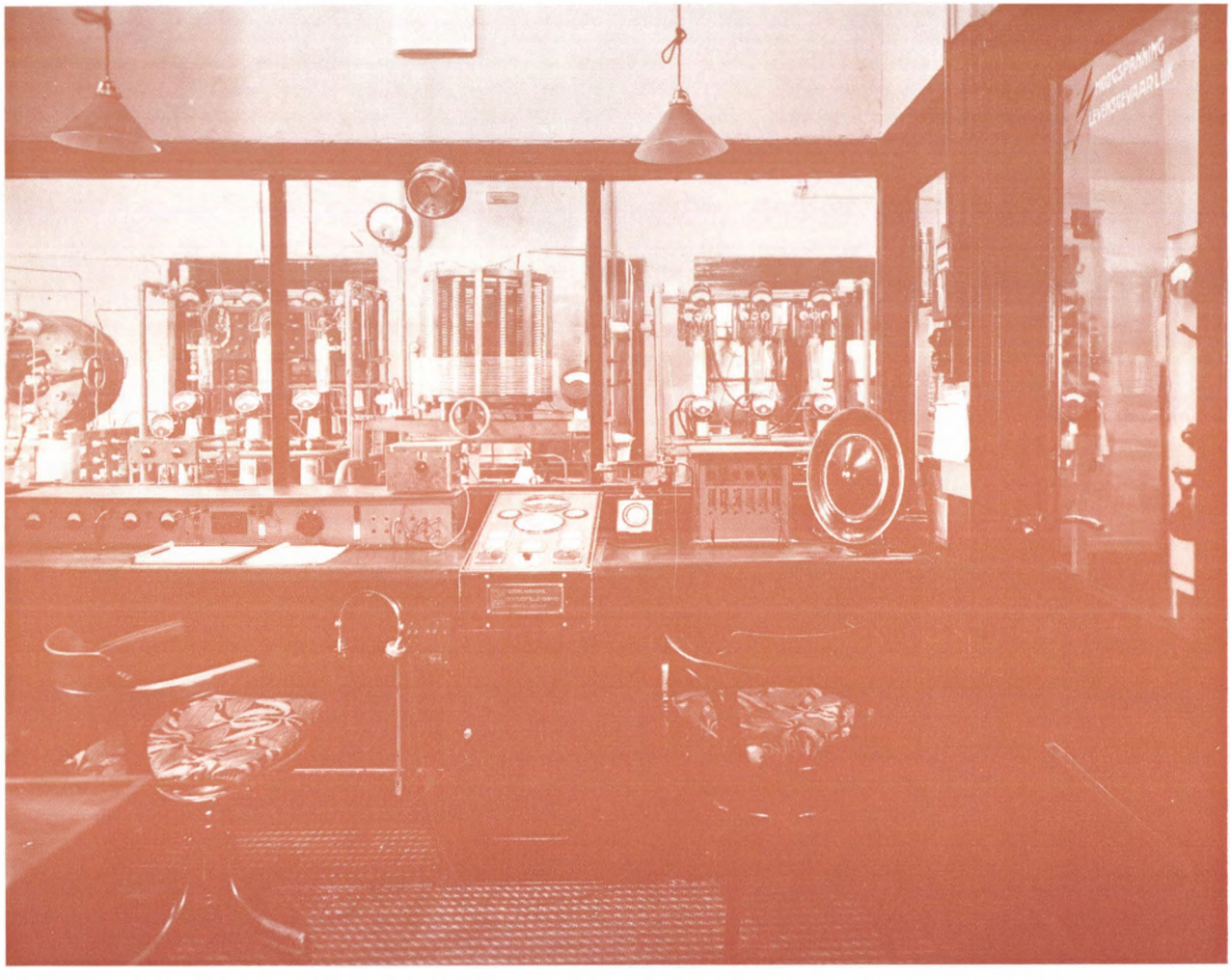


Fig. 15. Vector diagram for lower-sideband modulation with a sinusoidal information signal of angular frequency ω_p. The vector OS of the modulated signal $S(t)$ is constant in length, and rotates at an angular velocity $-\omega_p$. In upper-sideband modulation the angular velocity is ω_p.

There are also information signals for which even a constant phase difference cannot be tolerated. In television signals, for example, the steepness of pulse edges — transitions between dark and light — are important, and the shape of a pulse edge is very sensitive to the relative phase relations of the Fourier components. In such cases it is essential to have a fixed phase relationship between the reference signal and the carrier, and a pilot signal is therefore indispensable.

[2] A decibel (dB) is a logarithmic measure of power ratios. When the ratio is x the difference is said to be (10 log x) dB. A factor of 10 is a difference of 10 dB, a factor of 2 is very close to 3 dB.



An AM broadcasting transmitter built by Nederlandsche Seintoestellenfabriek in 1925. The transmitter operated at a wavelength of 1050 metres and radiated a power of 500 watts.

It will be evident that in two-way radio communication, where intelligible speech is the main requirement, while it is very important to keep the equipment simple, SSB modulation has great advantages over conventional AM and DSB modulation. These may be summarized as follows. In the first place the required bandwidth is equal to that of the basic signal, i.e. at least twice as small as in DSB modulation and conventional AM. In addition the peak power required for a given signal-to-noise ratio is another 3 dB lower than in DSB modulation — i.e. 9 dB lower than in conventional AM — because the detector only receives half as much noise, as a result of the halved receiver bandwidth.

If speech only is to be transmitted, the advantage compared with conventional AM is in one respect even greater. For the equipment it is often the mean power, not the peak power that is the important quantity; the lower the mean power the more compact the equipment can be. Now with a speech signal the mean power is very much lower than the peak power (of the order of 10 dB). Consequently the mean power transmitted in SSB and DSB modulation is 10 dB lower than the peak power. In conventional AM, on the other hand, this difference is no more than 6 dB; see *fig. 16*. The net result is thus an advantage of $9 + (10 - 6)$ dB $= 13$ dB in the mean power required when SSB modulation is compared with conventional AM. In such cases a 100-W SSB transmitter is therefore as effective as a 2-kW AM transmitter.

Generation of an SSB signal

As already indicated, an SSB signal can be obtained by first producing a DSB signal and then filtering out one sideband. This method is in fact used. Since, however, the two bands are close together, very good filter performance is required. We shall mention here two other widely used methods, which avoid the use of an r.f. filter.

A block diagram of the first of these methods is shown in *fig. 17*. It is based on the relation:

$$\cos \omega_p t \cos \omega_c t \pm \sin \omega_p t \sin \omega_c t = \cos (\omega_c \mp \omega_p) t. \quad (6)$$

The information signal $p(t)$ and the carrier $C(t)$ are multiplied together at two places: directly in M_1, and in M_2 after all the components have been shifted by 90° in phase in P_1 and P_2. M_1 and M_2 are modulators of the type shown in fig. 12. The two multiplications yield two DSB signals, i.e. the first and second terms in (6), which, added together or subtracted from each other, give the desired SSB signal.

This method is in practice only used for speech signals, that is to say signals with a relatively narrow frequency band. For these signals the required 90° phase shift of *each* Fourier component — the main problem of the method — is still feasible, although the 'ordinary' circuits used for this (with resistors, inductors and capacitors) are already fairly complicated. The 'transversal filters' discussed later (see p. 332) are a modern alternative for such circuits.

The operation that P_1 performs in the ideal case is known as a Hilbert transformation. The Hilbert transform $\tilde{p}(t)$ of a function $p(t)$ is:

$$\tilde{p}(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{p(\tau)}{t - \tau} \, d\tau. \quad (7)$$

It can easily be shown from this definition that the transformations of $\cos(\omega t + \phi)$ and $\sin(\omega t + \phi)$ give $\sin(\omega t + \phi)$ and $-\cos(\omega t + \phi)$. Since in addition the relation (7) between $p(t)$ and $\tilde{p}(t)$ is linear and additive, the transformation is in fact equivalent to the operation of P_1: the 90° rotation of each Fourier component. This notation allows the SSB signals of the information signal $p(t)$ to be represented in a closed form:

upper-sideband signal: $p(t) \cos \omega_c t - \tilde{p}(t) \sin \omega_c t$,
lower-sideband signal: $p(t) \cos \omega_c t + \tilde{p}(t) \sin \omega_c t$.

The second method we shall discuss here, generally known as the Weaver method, is illustrated in the block

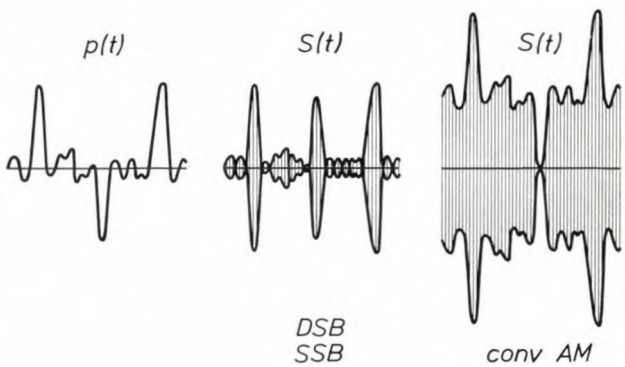


Fig. 16. In speech the average power of the information signal $p(t)$ is roughly 10 dB lower than the peak power. In DSB and SSB modulation this relationship is directly reflected in the modulated signal $S(t)$. In conventional AM with 100% modulation the average amplitude is however about half the peak amplitude, so that the mean power is only 6 dB lower than the peak power.

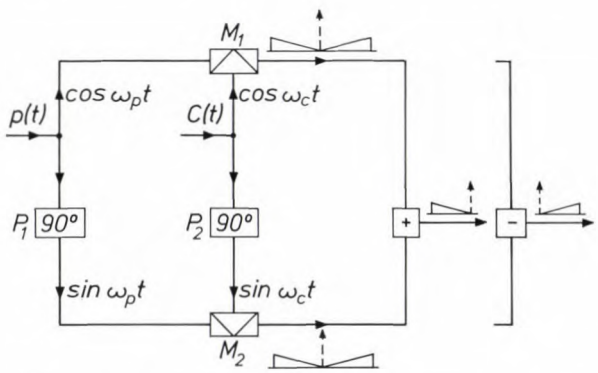


Fig. 17. Block diagram of a commonly used circuit for generating an SSB signal, which does not require an r.f. filter; $p(t)$ information signal, $C(t)$ carrier. M_1, M_2 multipliers. P_1, P_2 90° phase-shifters. Summation of the double-sideband signals from M_1 and M_2 yields the lower-sideband signal, subtraction yields the upper-sideband signal.

diagram in *fig. 18* [3]. Here the information signal is multiplied by $\cos \omega_0 t$ in M_1 and by $\sin \omega_0 t$ in M_1', where ω_0 is a *low* frequency, in the centre of the frequency band of the information signal. Signals at the sum frequency are eliminated from the products by filters F and F'. The remaining signals are multiplied in M_2 and M_2' by high-frequency signals $\cos(\omega_c + \omega_0)t$ and $\sin(\omega_c + \omega_0)t$. Addition of the two products thus obtained gives the upper-sideband signal (frequency $\omega_c + \omega_p$). To obtain the lower-sideband signal, the generator G_2 must deliver the frequency $\omega_c - \omega_0$, and the difference of the products obtained with M_2 and M_2' must be taken. This method does not require r.f. filters and the components of $p(t)$ do not have to be altered in phase.

In both methods the special combination of two DSB signals completely eliminates one of the sidebands in the ideal case. In practice a difference of 30 dB between the wanted and the unwanted sidebands is readily achieved. Often the resultant signal is passed to a final simple filter to give even greater suppression of the unwanted band.

Electronic problems such as those encountered here can nowadays often be solved more readily by digital processing of the signals.

Frequency-division multiplex (FDM)

Single-sideband modulation is very widely used today in telephony. In the 1920s the rapid growth of telephone communication created a need for a system in which different speech signals could be transmitted simultaneously along a single cable. This led to the development of the frequency-division multiplex system, FDM. In this system a number of speech signals modulate separate carriers in SSB in such a way that the modulated signals appear in adjacent separate bands of the frequency spectrum. They can then be transmitted over a common cable, and separated from each other at the receiving end by means of filters [4].

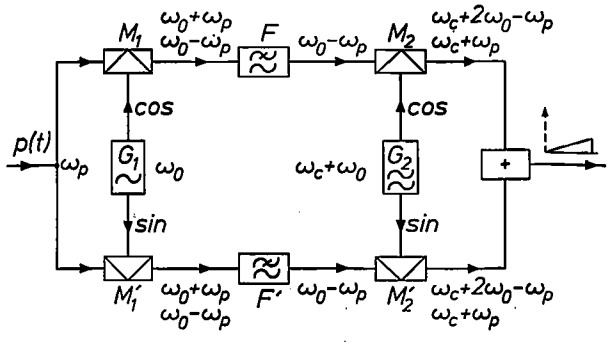


Fig. 18. The Weaver method used for generating an SSB signal. This avoids the use of r.f. filters and also a 90° phase shifter for the components of the information signal $p(t)$. G_1, G_2 are the sine and cosine generators of $\omega_0 t$ and $(\omega_c + \omega_0)t$ respectively. Here ω_0 is the centre of the baseband. M_1, M_1', M_2, M_2' multipliers. F, F' filters. Addition of the signals from M_2 and M_2' yields the upper-sideband signal. The lower-sideband signal is obtained by causing G_2 to generate the frequency $\omega_c - \omega_0$ instead of $\omega_c + \omega_0$ and taking the difference of the signals from M_2 and M_2'.

The carrier frequencies are multiples of 4 kHz. The smallest FDM group standardized by the CCITT (Comité Consultatif International de Télégraphie et Téléphonie) consists of 12 channels, which are *lower* sidebands (LSB) at the carrier frequencies 64, 68, 72, ... 108 kHz. The complete band thus goes from 60 to 108 kHz and is 48 kHz wide ('basic group'). To handle large volumes of telephone traffic several basic groups with LSB each in turn modulate a separate carrier. A 'supergroup' consists of five basic groups (60 channels) at carrier frequencies of 420, 468, ... 612 kHz. The lowest basic group is thus located in the band from 312 kHz (420 — 108) to 360 kHz (420 — 60), the second is in the band from 360 to 408 kHz, and so on. The total band goes from 312 to 552 kHz and is 240 kHz wide. There are also 'master groups' (five 'supergroups' consisting of 300 channels from 812 to 2044 kHz) and 'supermaster groups' (three 'master groups' consisting of 900 channels from 8516 to 12 388 kHz). Other groupings are also in use.

With such wide frequency bands it is necessary to take into account the frequency dependence of the attenuation in the transmission channel. In order nevertheless to obtain a flat transmission characteristic for the whole of the frequency band in use, some bands have to be given more amplification at repeaters or at the receiving end than others ('equalization'). To be able to determine at any given moment what the equalization characteristic and the overall gain should be, a number of pilot frequencies are transmitted at varying spacings within the band. These might also be used for synchronous detection, but there is generally no need for this since detection with stable local oscillators gives adequate results.

The transmission of so many signals by frequency-division multiplexing over a single transmission path requires very good linearity from all the amplifiers (repeaters), modulators and other equipment if interference is to be avoided from signals at sum and difference frequencies (caused by 'intermodulation' and 'cross-modulation').

Related methods of amplitude modulation

In DSB modulation the frequency band used is at least twice as wide as is necessary for the transmission of a given baseband. To make more efficient use of available bandwidths it is possible, instead of eliminating one sideband as in SSB modulation, to transmit two independent information signals in the two sidebands. In 'ISB' (independent-sideband) modulation, SSB methods are used to transmit one information

signal in the upper sideband and the other in the lower sideband of the same carrier. The practical significance of using the same carrier is mainly that only one pilot signal and one reference oscillator are necessary to detect both signals. Separation of the two signals in the receiver, however, calls for a very selective filter. The separation can also be effected by reversing the circuit of fig. 18 (the Weaver method); the circuits must then be very accurately balanced. ISB modulation is sometimes used in fixed radio links (p. 357) for the simultaneous transmission of different speech or telegraph signals.

As we have seen on page 311, the output signal in synchronous detection is zero if the reference phase differs by 90° from that of the carrier. This gives yet another way of transmitting *two* signals in the sidebands of a single carrier. If for simplicity we take $\phi = 0$ in eq. (4), then our original modulated signal is $p(t) \cos \omega_c t$. We now add a signal that is 90° out of phase and is modulated by another information signal, $q(t)$:

$$S(t) = p(t) \cos \omega_c t + q(t) \sin \omega_c t.$$

This is shown by the vector diagram in *fig. 19*. The total signal now varies not only in amplitude, as before, but also in phase. As can easily be shown, synchronous detection with $\cos \omega_c t$ as reference signal yields the signal $p(t)$, while synchronous detection with $\sin \omega_c t$ as reference signal yields $q(t)$. Small phase errors, not serious with DSB modulation, now lead however to crosstalk between p and q.

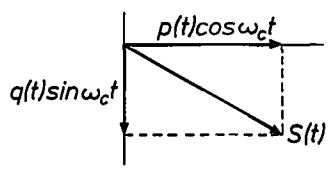


Fig. 19. Vector diagram for quadrature amplitude modulation QAM). As in DSB modulation, the signals $p(t) \cos \omega_c t$ and $q(t) \sin \omega_c t$ change only in length in the vector diagram and not in direction, but they change independently, so that the vector sum $S(t)$ generally changes both in length and direction.

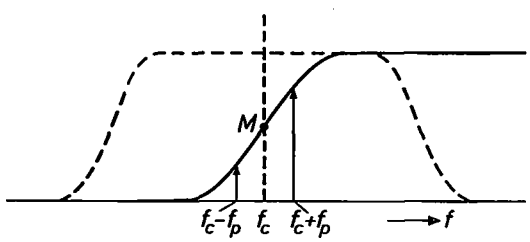


Fig. 20. Vestigial-sideband modulation (VSB). The VSB signal is a conventionally modulated signal (dashed line) which has passed through a filter whose amplitude-response characteristic is shown by the solid line. The filter transmission is 50% at f_c; the characteristic is symmetrical with respect to point M.

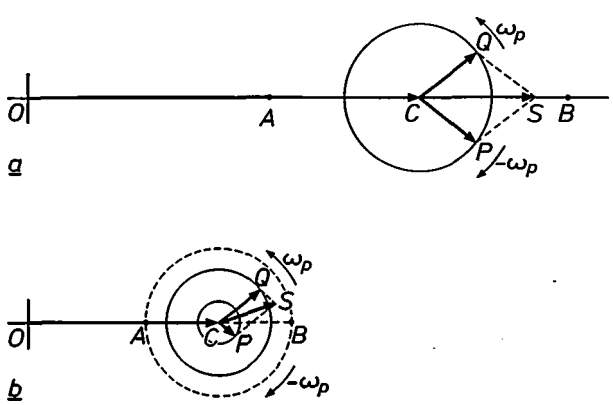


Fig. 21. Vector diagrams *a*) for conventional AM, *b*) for VSB modulation. Because of the symmetry of the filter in VSB modulation (fig. 20), the length of $CP + CQ$ in (*b*) is the same as CP in (*a*), and OC in (*b*) is half OC in (*a*). The relations between the variation AB in the length of the vector sum OS and the carrier amplitude OC are thus the same in VSB modulation as in conventional AM, so that peak detection produces much the same result.

Finally, there is another method, used in television and in other forms of communication, for modulating the main carrier. During the development of monochrome television it was decided — as in the case of sound broadcasting because of the large number of receivers — to adopt a method in which peak detection could be used. At the same time it was also desirable to be economical of bandwidth. This resulted in the *vestigial-sideband system* (VSB). In this system a conventionally modulated signal is passed through a filter with the amplitude-response characteristic shown in *fig. 20*.

From *fig. 21* it can easily be seen that peak detection now gives about the same results as in conventional amplitude modulation. Fig. 21*a* again shows the vector diagram for conventional AM; the two components CP and CQ from the lower and upper sidebands rotate in opposite senses at an angular velocity ω_p; the resultant CS varies only in amplitude, and peak detection yields the variation AB in amplitude of the total vector OS ($CA = CB = 2CP$). Fig. 21*b* is the vector diagram for VSB modulation. The resultant OS of the carrier OC and the components CP and CQ from the lower and upper sideband now vary in phase as well. Owing to the symmetry of the filter the sum of the lengths CP and CQ in (*b*), however, is equal to the length CP in (*a*), so that the detected variation AB in the amplitude of

[3] D. K. Weaver Jr., Proc. IRE 44, 1703, 1956.
[4] See for example H. N. Hansen, Philips tech. Rev. 26, 206; 1965.

OS is half that in (a). Since the carrier is also halved, the modulation depth remains unchanged. If the filter has a sharp cut-off, the vector CP in (b) is zero at higher f_p, and the vector point Q rotates around the dashed circle. The phase variations — negligible when the modulation depth is sufficiently small — result in some distortion. In television these variations are generally tolerable, although a correction is sometimes made at the transmitter to compensate for the distortion. However, synchronous detection is steadily replacing peak detection in television receivers, and this distortion is not encountered with synchronous detection; the reference signal is obtained by sharply filtering out the carrier.

Frequency and phase modulation

Amplitude modulation may be described as the substitution of a function of time, e.g. $C + p(t)$, for the constant A in the expression for the carrier

$$C(t) = A \cos(\omega_c t + \phi). \tag{8}$$

In frequency and phase modulation the situation is not so straightforward since variations in phase and variations in frequency are inseparably related [5]. This can be seen immediately if the constant ϕ in (8) is replaced by a function of time, $\phi(t)$. The result is a waveform of the type shown in *fig. 22*, and it can be seen that the frequency — considered as the number of periods of $S(t)$ per second — also fluctuates. To investigate this further we must first take a closer look at the concept of 'frequency' in this new situation.

A carrier of the type in fig. 22 can generally be represented by

$$S(t) = A \cos \psi(t), \tag{9}$$

where A is a constant and $d\psi/dt$ is greater than zero. When, in a certain time interval Δt, a whole number of periods of $S(t)$ have elapsed, this number divided by Δt is the 'mean frequency' over Δt. The number of periods elapsed is the number of times that ψ has increased by 2π, i.e. $\Delta\psi/2\pi$, and the mean frequency is thus $(1/2\pi)\Delta\psi/\Delta t$. This expression can also be used when the number of periods that have elapsed in Δt is not an integer. The value to which this mean frequency approaches for $\Delta t \to 0$ is called the *instantaneous frequency* f_m:

$$f_m = \lim_{\Delta t \to 0} \frac{1}{2\pi} \frac{\Delta\psi}{\Delta t} = \frac{1}{2\pi} \frac{d\psi}{dt}.$$

The corresponding *instantaneous angular frequency* ω_m is equal to $d\psi/dt$.

In the case of an unmodulated carrier the instantaneous angular frequency is of course equal to the angular frequency ω_c of the carrier. In the case of a fluctuating phase $\phi(t)$ we have:

$$\psi = \omega_c t + \phi(t), \tag{10}$$

$$\omega_m = \omega_c + d\phi/dt. \tag{11}$$

With a given information signal $p(t)$ the terms *phase modulation* and *frequency modulation* are now used as follows. If the phase $\phi(t)$ in (10) varies in proportion to $p(t)$, we have phase modulation. The instantaneous frequency then varies in proportion to dp/dt from (11). Conversely, we have frequency modulation when the deviations in the instantaneous frequency are proportional to $p(t)$; in that case $d\phi/dt$ is proportional to $p(t)$ and the phase thus varies in proportion to $\int p dt$.

Frequency modulation is therefore *not* the substitution of a variable such as $\omega_c + ap(t)$ for the constant ω_c in (8). This would lead to ridiculous results. If we assume for instance that $p(t)$ is periodic: $p(t) = \cos \omega_p t$, then we would find:

$$\psi = (\omega_c + a \cos \omega_p t)t + \phi = \omega_c t + at \cos \omega_p t + \phi. \tag{12}$$

Fig. 23 shows the variation of ψ as given by (12). This is a completely unusable form of modulation: the variation of ψ, and hence of S, eventually becomes too violent.

We shall now confine ourselves for the moment to the case in which the information signal is sinusoidal and the phase $\phi(t)$ is given by:

$$\phi(t) = \alpha \cos \omega_p t. \tag{13}$$

In the vector diagram (*fig. 24*) the vector of the modulated signal,

$$S(t) = A \cos(\omega_c t + \alpha \cos \omega_p t), \tag{14}$$

maintains a constant amplitude A, but the phase angle oscillates to and fro around the carrier vector. The angular amplitude α of this motion is known as the *modulation index*. The instantaneous angular frequency ω_m is:

$$\omega_m = \omega_c - \alpha\omega_p \sin \omega_p t.$$

The *frequency deviation* Δf, i.e. the amplitude of the variation in instantaneous frequency, is thus given by:

$$\Delta f = \alpha f_p. \tag{15}$$

Since the instantaneous frequency oscillates to and fro between the values $f_c + \Delta f$ and $f_c - \Delta f$, the obvious conclusion might seem to be that the required bandwidth is $2\Delta f$, and is thus directly proportional to the degree of modulation (which is equal to Δf or to α, depending on whether frequency or phase modulation is under consideration). Because of this it was originally thought that in those cases where a low degree of modulation was considered sufficient, frequency or phase modulation could be used to save bandwidth, since it

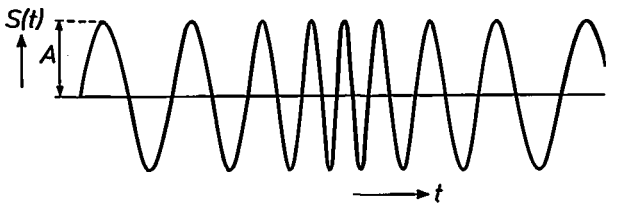


Fig. 22. Carrier modulated in phase and in frequency. Modulation of this type can be obtained by varying ϕ in (8) as a function of time. The duration of a complete period, and hence the frequency, then varies as well.

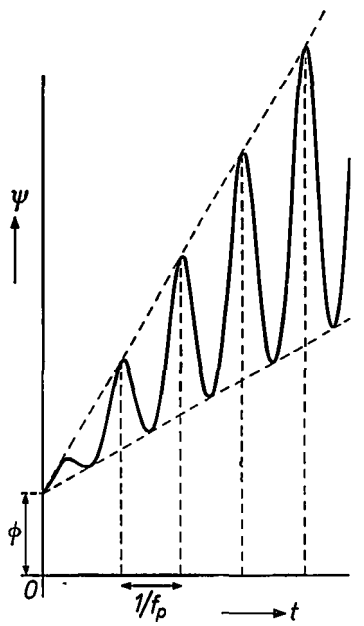


Fig. 23. The argument $\psi(t) = \omega_c t + \phi$ of $S(t)$ (see eq. 9), as a function of t, when ω_c is replaced by a constant plus a periodic function.

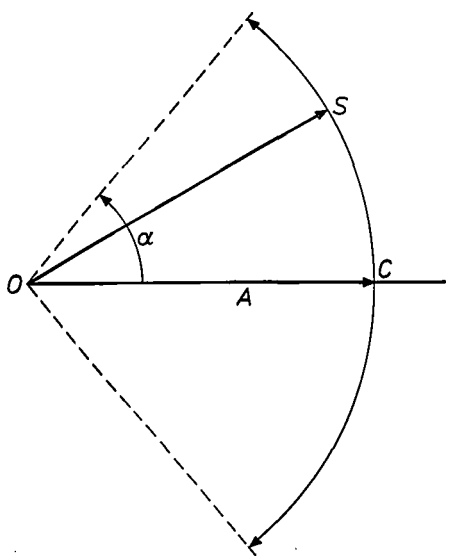


Fig. 24. Vector diagram for frequency or phase modulation. OC vector of the unmodulated carrier. OS, the vector of the modulated signal, does not change in length but its direction varies periodically. The angular amplitude α of this movement is the 'modulation index'.

follows from (15) that $2\Delta f$ is smaller than the $2f_p$ required by conventional AM, if α is less than 1. This conclusion is *not* correct; as we shall presently see, FM always requires a frequency band at least as wide as that for AM.

The usefulness of FM does not therefore lie in any saving of bandwidth, but on the contrary in the possibility it offers of using an α *greater* than 1. At a given carrier amplitude A, that is to say at a given transmitter power, the modulation amplitude in conventional AM is limited to $\frac{1}{2}A$, and in DSB and SSB modulation to A. From equations (13) and (14) and fig. 24 it is evident that in FM the admissible degree of modulation (α) is not limited by the transmitter power in such a simple way. This means that at a given transmitter power a much better signal-to-noise ratio can generally be obtained than with AM. This does however use up a bandwidth of *at least* $2\Delta f$, which increases with the α-value selected. We shall presently examine in more detail what the bandwidth actually is and the extent to which the signal-to-noise ratio can be improved compared with AM.

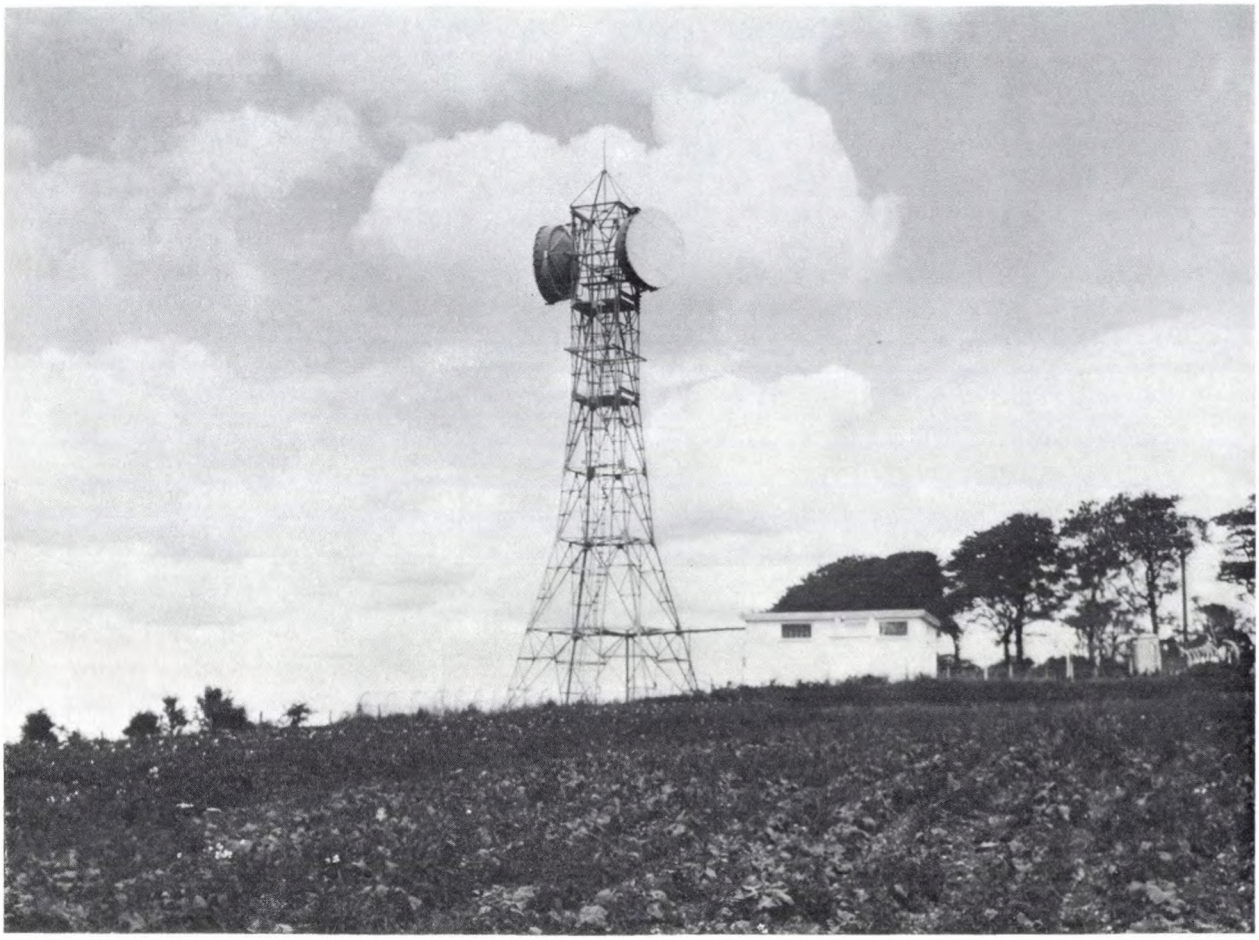
The great virtue of FM really lies in the fact that the zero crossings of the signal contain all the information. Small *additive* interferences such as noise — small vertical displacements of $S(t)$ in fig. 22 — cause only slight shifts in the zero crossings; those due to the signal can be made much greater by choosing a sufficiently large α. *Multiplicative* interference (amplitude variations) have no effect at all on the zero crossings, at least as long as the amplitude does not fall to zero. This also applies at small values of α. In FM, therefore, the information is well protected from unintentional amplitude variations. This is the second important advantage of FM.

Modulation and detection

To modulate a carrier in frequency or in phase it is only necessary to vary the capacitance of the frequency-determining LC circuit in the oscillator or of a phase-shifting filter after the oscillator. This is what is done in practice, with the aid of solid-state 'varactor diodes', which have a voltage-sensitive capacitance. Because of the insensitivity of FM to amplitude errors, there are no special requirements for the linearity of the output amplifier. For this reason an FM transmitter can be simpler than an AM transmitter — an important advantage in mobile installations.

The detector of an FM signal must deliver a current or voltage that follows the variations in the instantaneous frequency. Many methods have been developed for this, two of which will be mentioned here. In the first

[5] See for example Balth. van der Pol, Proc. IRE 18, 1194, 1930.



Relay station in a microwave link. Traffic is sent simultaneously in both directions. The information is sent on a frequency-modulated carrier; the wavelength is between 3 and 15 centimetres. The parabolic reflectors direct the signals into a beam with an aperture of about 1°. A link of this kind can carry up to 2700 simultaneous telephone conversations per carrier in frequency-division multiplex (FDM). The capacity can be increased by using more than one carrier.

method, the signal is directly applied to a circuit that converts the frequency variations into amplitude variations. The simplest example is a tuned circuit, with the side of the resonance curve covering the required frequency range; see *fig. 25*. The resultant AM signal can be demodulated by a peak detector. FM detection does not therefore need to be much more complicated than ordinary AM detection. To ensure linear detection, however, the amplitude-frequency response should be sufficiently linear. In this respect the side of a tuned-circuit response is not the ideal solution. More refined circuits, like the Foster-Seeley detector and the ratio detector, use coupled tuned circuits. An additional advantage of the ratio detector is that it is relatively insensitive to amplitude variations in the input signal.

In the second method the zero crossings of the signal are converted into electrical pulses of constant shape and amplitude, and the pulse train thus obtained is then electronically integrated over short periods. The resultant signal is proportional to the instantaneous frequency.

In all FM detection methods, amplitude variations are removed before detection by clipping the signal peaks in an amplitude limiter.

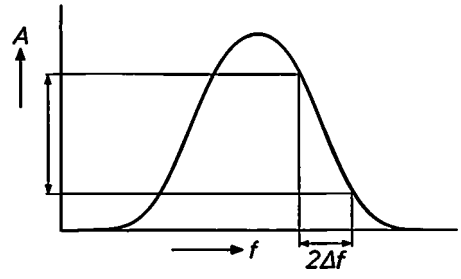


Fig. 25. Amplitude-frequency characteristic of a tuned circuit. The side of the response curve can be used in frequency demodulation to convert frequency variations into amplitude variations.

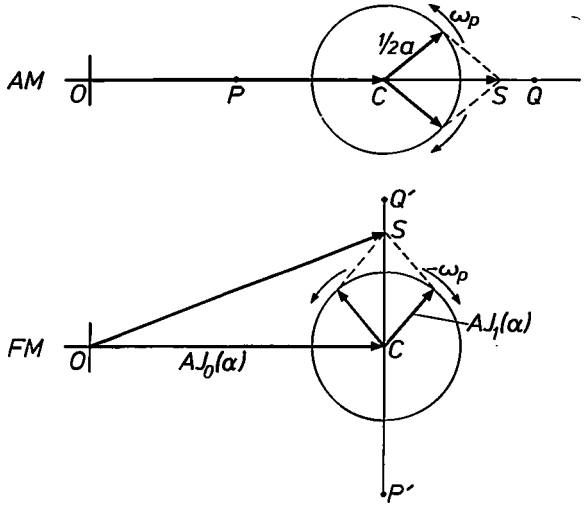


Fig. 26. Vector diagrams for conventional AM and for FM taking into account only the first-order sidebands. In both cases the vectors of the sideband signals rotate in opposite directions at an angular velocity ω_p. The point S corresponding to the vector sum OS moves to and fro along the line of the carrier vector OC in AM, but in this first-order approximation it moves perpendicular to OC in FM.

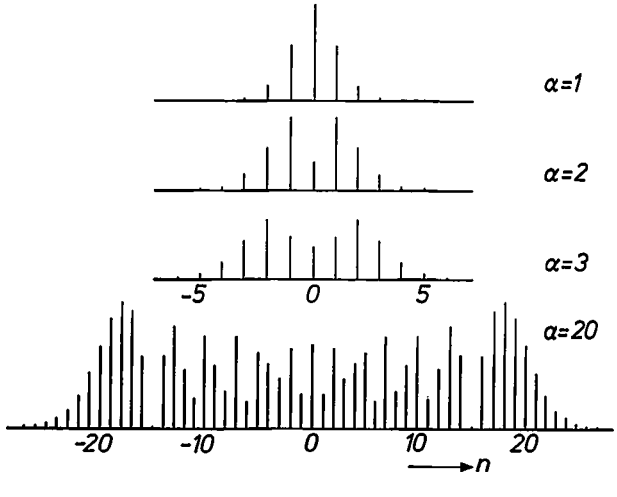


Fig. 27. Spectra of signals modulated in frequency or in phase by a sinusoidal information signal; α is the modulation index. The amplitude of the component of frequency $f_c \pm nf_p$ is proportional to $|J_n(\alpha)|$, the absolute value of the Bessel function of the nth order of α.

Bandwidth

When the instantaneous frequency f_m changes slowly at a given frequency deviation Δf, in other words when the signal frequency f_p is small and the modulation index α therefore large, the instantaneous frequency is 'quasi-steady-state'. To a good approximation, we are then only concerned with the frequencies that lie between the maximum and the minimum of f_m, so that the bandwidth is indeed $2\Delta f$. For small values of α ($\alpha < 1$) this is certainly not the case. To determine the bandwidth correctly we have to expand the modulated signal (14) as a Fourier series. In principle the result is an *infinitely broad spectrum*, because there is a component for every angular frequency $\omega_c \pm n\omega_p$ (n integer). Its amplitude is equal to the absolute value of the Bessel function of the nth order of α, $|J_n(\alpha)|$.

This result is found by writing the modulated signal (14) in the form

$$S(t) = A\cos\omega_c t\cos(\alpha\cos\omega_p t) - A\sin\omega_c t\sin(\alpha\cos\omega_p t),$$

substituting in this the following series for $\cos(\alpha\cos\omega_p t)$ and $\sin(\alpha\cos\omega_p t)$, the Jacobi series:

$$\cos(\alpha\cos\omega_p t) = J_0(\alpha) - 2J_2(\alpha)\cos 2\omega_p t + 2J_4(\alpha)\cos 4\omega_p t \ldots,$$
$$\sin(\alpha\cos\omega_p t) = 2J_1(\alpha)\cos\omega_p t - 2J_3(\alpha)\cos 3\omega_p t + \ldots$$

and then replacing each product of $\cos\omega_c t$ or $\sin\omega_c t$ and $\cos n\omega_p t$ by a term at the sum frequency and a term at the difference frequency. A term with a negative sign can be taken care of by an extra phase term of 180°.

If the Fourier series is discontinued after the terms of the first order, we obtain a simple parallel with AM. The result obtained and the expression for AM (eq. 3) with $\phi = 0$ are respectively:

$$S(t) = AJ_0(\alpha)\cos\omega_c t - AJ_1(\alpha)\sin(\omega_c + \omega_p)t - AJ_1(\alpha)\sin(\omega_c - \omega_p)t, \quad \text{(FM)}$$

$$S(t) = C\cos\omega_c t + \tfrac{1}{2}a\cos(\omega_c + \omega_p)t + \tfrac{1}{2}a\cos(\omega_c - \omega_p)t. \quad \text{(AM)}$$

Fig. 26 shows the corresponding vector diagrams. Because of the differences in the phases of the sideband components, the resultant of the vector sum in FM moves up and down along the vertical line $P'Q'$ and not along the horizontal line PQ as in AM. In the first case we have mainly a rotation of the vector sum, and in the second a change in length. When the terms of higher order are included, the straight line $P'Q'$ becomes a circular arc (fig. 24).

Fig. 27 shows the spectra for $\alpha = 1$, $\alpha = 2$, $\alpha = 3$ and $\alpha = 20$. These illustrate two important facts. In the first place, at a given α the components for which n exceeds a particular value can be neglected. In the second place, the larger the value of α the more components have to be taken into account. Neglecting components implies signal distortion. It is therefore necessary to compromise between a strong signal (implying strong modulation), a narrow bandwidth and low distortion. *Fig. 28* gives curves of α as a function of n for several values of the distortion δ. This quantity is defined by the rule that all components in the spectrum whose amplitude $|J_n(\alpha)|$ is less than δ are suppressed.

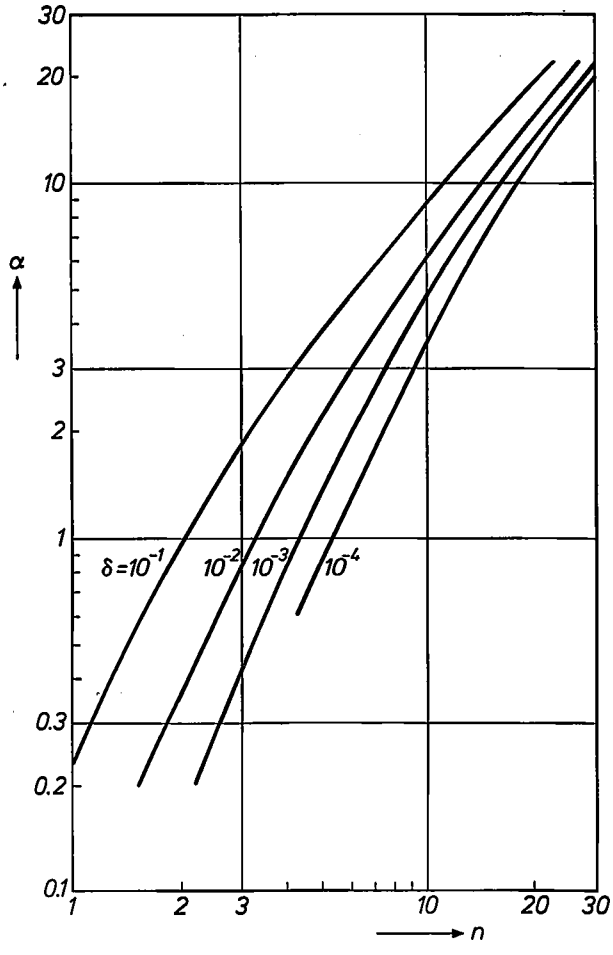


Fig. 28. α as a function of n for some values of $\delta = |J_n(\alpha)|$. This indicates how large the frequency deviation Δf ($= \alpha f_p$) at a given bandwidth B ($= 2nf_p$) can be before the distortion δ exceeds a particular limit. At a distortion δ all Fourier components are cut off whose amplitude is smaller than δ times that of the unmodulated carrier.

Fig. 28 shows, for example, that at a permitted distortion of 10^{-3}, three components on either side of the carrier are sufficient for a modulation index of $\alpha = 0.2$, whereas as many as seven are required for $\alpha = 2$.

For small values of α ($\alpha \ll 1$) all components can be disregarded except for those of the zero and first orders ($n = 0$ and 1). The bandwidth is then equal to that for conventional AM, i.e. $2f_p$. FM therefore requires at least as much bandwidth as AM, and in general much more.

The spectrum for $\alpha = 20$ in fig. 27 illustrates the fact that at large values of α the components can soon be disregarded when n becomes greater than α. The bandwidth is then approximately equal to $2\alpha f_p$, and therefore approximately equal to twice the frequency deviation (see eq. 15), in accordance with the prediction for this quasi-steady-state situation.

Summarizing, the bandwidth B is equal to $2f_p$ for small α, and to $2\Delta f$ for large α. In intermediate cases the usual rule of thumb is:

$$B = 2(\Delta f + f_p).$$

So far we have only considered a sinusoidal signal. If the information signal is built up from a large number of Fourier components, the FM signal, unlike an AM signal, cannot be obtained by superposition of the FM signals of the components ('FM is not linear'). Sidebands also occur at all possible combinations of frequencies. In this case, however, there is again a very useful rule of thumb for the bandwidth, known as Carson's rule:

$$B = 2(\Delta f + f_{max}). \qquad (16)$$

Here f_{max} is the maximum frequency of the baseband and Δf is the frequency deviation of the total modulated signal (i.e. $1/2\pi$ times the maximum of $|d\phi/dt|$ in (11)).

Signal-to-noise ratio

We shall first consider the situation where a carrier (amplitude A, frequency f_c) enters the detector together with a small sinusoidal interfering signal of amplitude b and frequency $f_c + f_n$, where f_n is a low frequency. In the vector diagram (*fig. 29*) the interfering vector $\boldsymbol{b}$ rotates at an angular velocity ω_n. For phase detection the interference corresponds to a signal of frequency f_n from the baseband and modulation index α_n. We assume that b is much smaller than A; then $\alpha_n = b/A$. If in addition the carrier is phase-modulated by a signal of frequency f_p and modulation index α, the ratio of the signal to the interference is:

$$\alpha/\alpha_n = \alpha A/b.$$

Now A/b happens to be equal to the signal-to-noise ratio in AM with 100% modulation depth. After phase detection the signal-to-noise ratio is thus α times greater than in AM.

In frequency detection the interference corresponds to a signal with a frequency deviation Δf_n, given by

$$\Delta f_n = \alpha_n f_n. \qquad (17)$$

After detection the ratio of the signal Δf to the interference Δf_n is therefore:

$$\frac{\Delta f}{\Delta f_n} = \frac{\alpha f_p}{\alpha_n f_n} = \left(\frac{\alpha f_p}{f_n}\right)\frac{A}{b}.$$

In this case the factor of improvement compared with AM is thus greater the closer the interfering frequency $f_c + f_n$ lies to f_c. The reason for this is that the frequency excursions Δf_n of the interference are proportional to f_n (see 17). This also implies that if the FM transmission channel contains *white* noise — a flat spectrum of uncorrelated interference vectors (b and α_n independent of f_n) — the low-frequency amplitude-noise spectrum after detection will be peaked ('trian-

gular noise'); the power spectrum is then parabolic (*fig. 30*).

We now consider an information signal of frequency f_p from a baseband 0-f_{max}, and an FM transmission channel of bandwidth B (*fig. 31*). We let the frequency deviation ΔF of the signal be as large as the bandwidth allows. For simplicity we assume that B is much larger than f_{max}, so that $\Delta F \approx \frac{1}{2}B$, and thus independent of f_p. After detection the signal contains low-frequency noise originating from high-frequency noise from the whole band B. A filter eliminates the low-frequency noise at frequencies above f_{max}. For the case of white noise in the FM channel we now calculate the ratio of the detected (low-frequency) noise power (N) in the baseband to the detected signal power (S). The low-frequency noise power is the sum of contributions from the subbands df_n. We put $b^2 = n df_n$; apart from a constant factor, n is the noise power per Hz of bandwidth before the detector. We then find:

$$\frac{N}{S} = \frac{\sum(\Delta f_n)^2}{(\Delta F)^2} = \frac{1}{(\Delta F)^2}\sum\frac{b^2}{A^2}f_n^2 = \frac{n}{(A\Delta F)^2}\int_0^{f_{max}} f_n^2 df_n = \frac{n f_{max}^3}{3(A\Delta F)^2}.$$

For AM detection we would have:

$$\frac{N}{S} = \frac{\sum b^2}{A^2} = \frac{1}{A^2}\int_0^{f_{max}} n df_n = \frac{n f_{max}}{A^2}.$$

The improvement factor in the ratio of the signal power to the noise power, S/N, in FM as compared with AM is thus $3(\Delta F/f_{max})^2 = 3\alpha_0^2$. Here α_0 is the permissible modulation index at the highest frequency of the information signal. If, for example, α_0 is equal to 5, then S/N in FM is 75 times greater than in AM.

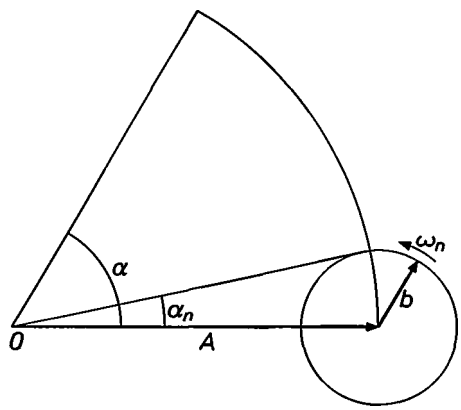


Fig. 29. Vector diagram for phase or frequency modulation, with an interfering signal $b \cos(\omega_c + \omega_n)t$ at the detector. For phase detection the interference corresponds to a signal of modulation index α_n.

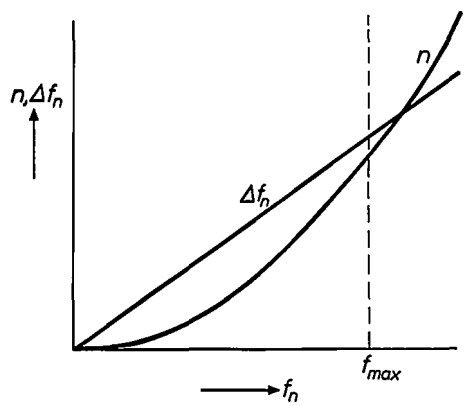


Fig. 30. 'Triangular noise'. In frequency detection the frequency deviation $\Delta f_n = \alpha_n f_n$ of an interfering signal is detected. White noise before the detector can be regarded as uncorrelated interfering vectors b (see fig. 29), where b, and hence α_n, are independent of f_n. The low-frequency amplitude-noise signal (Δf_n) at the output of the FM detector is thus proportional to f_n, and the noise-power spectrum (n) is proportional to f_n^2.

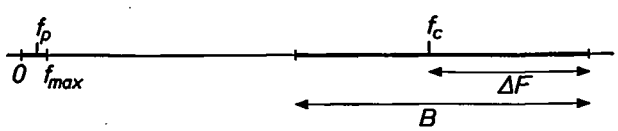


Fig. 31. Baseband 0-f_{max} and transmission channel of bandwidth B. Assuming that B is much larger than f_{max}, the maximum frequency deviation ΔF is very nearly equal to $\frac{1}{2}B$.

Information theory shows that the signal-to-noise ratio after the detector, $(S/N)_o$, can be improved by increasing the bandwidth of the channel combined with an appropriate signal treatment. The theoretically possible improvement, for not unduly small values of $(S/N)_i$, is given by the relation

$$(S/N)_o \approx (S/N)_i^{B/f_{max}}.$$

Here $(S/N)_i$ is the ratio of the signal power in the channel to the noise power generated in the channel. In FM the improvement is proportional to $(B/f_{max})^2$, as has just been demonstrated. Thus, FM gives an improvement of S/N when a larger bandwidth is used, but the improvement is far from what is theoretically possible.

Pre-emphasis

We assume now that the information signal $p(t)$ is not a pure sine wave but nearly sinusoidal, and that the frequency f_p varies slowly to and fro in the baseband 0-f_{max} with unchanging amplitude (*fig. 32a*). With phase modulation (fig. 32*b*), α is then constant and Δf is proportional to f_p. The highest value of Δf, at f_{max}, must not exceed the maximum ΔF set by the bandwidth of the channel. At lower frequencies Δf is smaller and optimum use is therefore not made of the channel bandwidth. Frequency modulation (fig. 32*c*) does make optimum use of the channel bandwidth if Δf — which here is everywhere the same — is made equal to ΔF.

With normal audio signals, e.g. music, the power decreases in general with rising frequency. To make

optimum use of the transmission channel over the whole baseband in frequency modulation, the signals at the higher frequencies are amplified with respect to those with lower frequencies before modulation (pre-emphasis), in such a way that the spectrum of the new information signal $q(t)$ becomes approximately flat. After modulation the reverse operation (de-emphasis) is applied to restore the original relationship between the components. *Fig. 33* shows some simple examples of pre-emphasis and de-emphasis circuits used in FM radio, together with the pre-emphasis obtained. At high frequencies the pre-emphasis is proportional to f_p; at the higher frequencies it therefore really amounts to transmitting $p(t)$ with phase modulation (see fig. 32*b*). In this way an additional improvement of the signal-to-noise ratio is obtained.

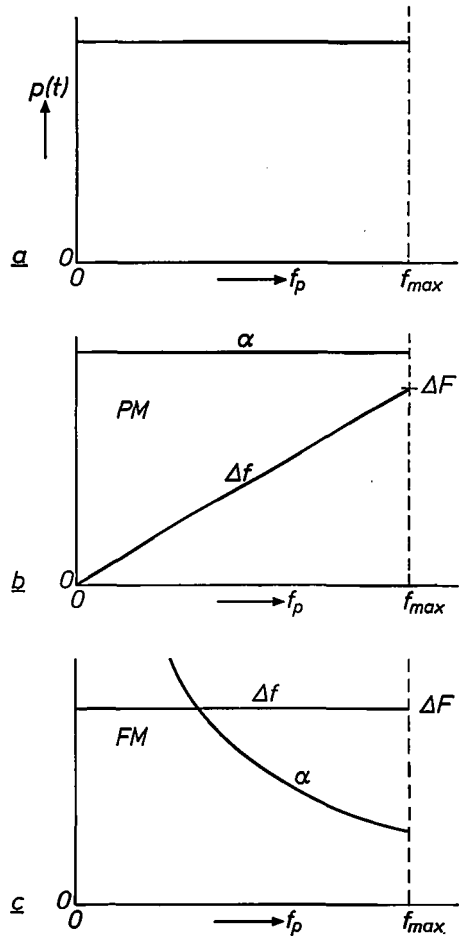


Fig. 32. When the information signal varies in frequency while maintaining a constant amplitude (*a*), α is constant and Δf proportional to f_p in phase modulation (*b*), whereas Δf is constant and α inversely proportional to f_p in frequency modulation (*c*). Phase modulation does not therefore make optimum use of the bandwidth ($2\Delta F$) of the transmission channel at all frequencies, whereas FM does.

Threshold effect

As long as the amplitude b of the interfering signal is smaller than A in fig. 29, the phase variations it produces are limited to the interval $\pm \pi/2$; the resultant oscillates backwards and forwards in the vector diagram. The situation becomes entirely different as soon as the interference becomes stronger than the signal ($b > A$); the resultant then *rotates* [6]. Whereas in the first case the zero in the signal $S(t)$ only shifted backwards and forwards, in the second case the signal $S(t)$ completes one period more or less than the carrier at every rotation of the resultant, so that there is one extra or one less zero. In the first case the phase of the total signal approximately follows the phase of the original signal; in the second case it follows that of the interference signal. Within the detector band it is thus the stronger of two FM signals that is received; the weaker, distorted beyond recognition, is audible as noise. If the strength ratio changes, there is only a narrow transition region in which both signals are unrecognizable. In mobile telephone communication, for example, this is an important consideration (p. 360).

In the case of one signal, distorted by noise, this effect sets a limit to the usefulness of increasing the frequency deviation and the bandwidth. If the bandwidth is increased, so also is the noise power at the detector. The vector diagram in *fig. 34* shows the noise of the detector schematically as the area of a circle of

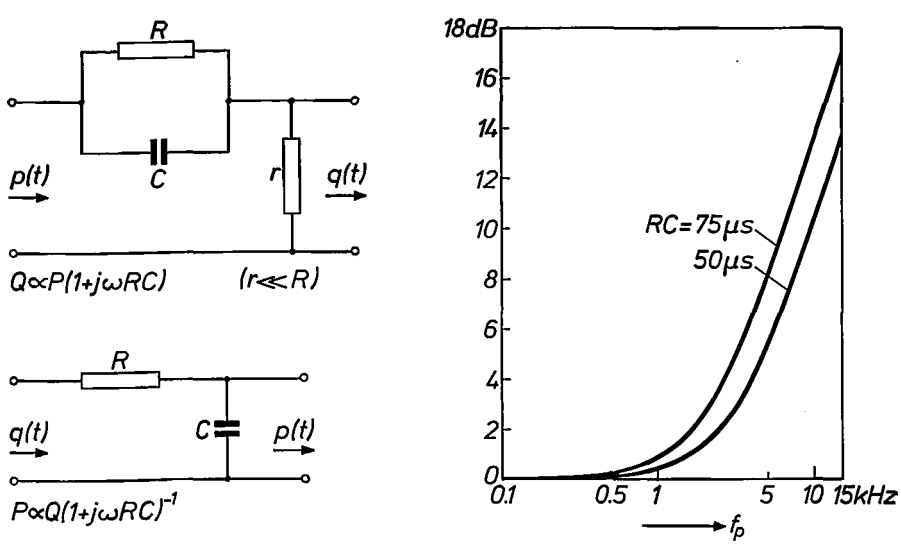


Fig. 33. Pre-emphasis and de-emphasis circuits and the pre-emphasis obtained for two values of the product RC. The quantities P and Q are the complex amplitudes of $p(t)$ and $q(t)$ as a function of ω.

radius b. When the modulation index α is increased, the growth of the bandwidth results in a larger 'noise circle'. When b approaches the carrier amplitude A, the threshold is reached at which any further increase in frequency deviation has an adverse effect.

The situation is summarized in *fig. 35*. For FM and AM (with synchronous detection) the figure gives the signal-to-noise power ratio $(S/N)_o$ after detection, in a log-log plot as a function of $(S_c/N)_i$, which is the ratio, before detection, of the carrier power to the noise power in a bandwidth equal to that of the baseband signal. In AM (curve AM) the two quantities are always identical ($A^2/\Sigma b^2$ in fig. 34). At not too small values of $(S_c/N)_i$ the value of $(S/N)_o$ in FM (curve FM_1) is higher than in AM by a fixed factor (a number of decibels that is independent of $(S_c/N)_i$). If the distance from the transmitter to the receiver is now increased, $(S_c/N)_i$ decreases because A in fig. 34 becomes smaller and therefore b becomes relatively larger. When b approaches A, the advantage of FM (after the threshold D_1) is reduced. When the threshold D_2 is reached, FM in fact becomes worse than AM. In practice D_1 would be 10 to 12 dB and D_2 6 to 8 dB. If the frequency deviation and the bandwidth are increased in FM, the gain with FM is greater (curve FM_2), but on decreasing $(S_c/N)_i$ the thresholds are reached earlier because the noise circle was initially larger.

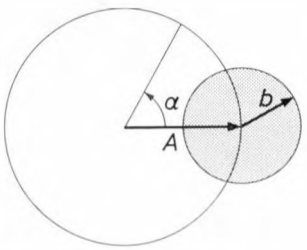


Fig. 34. Threshold effect. If α and hence the frequency deviation and bandwidth are increased to improve the signal-to-noise ratio, the total noise power at the detector also increases. This implies that the 'noise circle' (radius b) becomes larger. If b is no longer small compared with A, the threshold has been reached at which a further increase of α leads to a worse result.

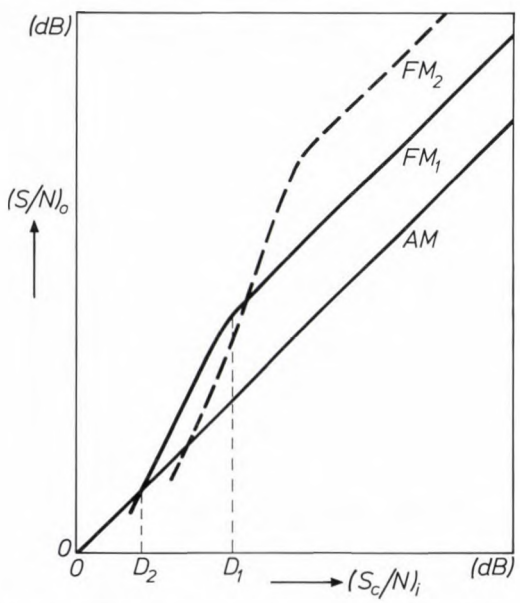


Fig. 35. Ratio of low-frequency signal power to low-frequency noise power after the detector, $(S/N)_o$, as a function of the ratio of the carrier power to the r.f. noise power in a band with the same bandwidth as the baseband signal at the detector, $(S_c/N)_i$. In AM with synchronous detection both quantities are identical. For large values of $(S_c/N)_i$, FM is better than AM by a constant number of decibels; at the threshold D_1 the advantage diminishes, and below D_2, FM is worse than AM (curve FM_1). With a larger frequency deviation (and bandwidth) the curve for large $(S_c/N)_i$ shifts upward (FM_2), but the thresholds shift to the right.

Threshold extension

It may sometimes be necessary to work with weak signals, for example in reception of signals from a communication satellite owing to the very limited power that the satellite can deliver, or in communication between distant stations that lie in each other's Earth shadow, or when the communication is brought about by the scattering of radio signals from inhomogeneities in the troposphere or ionosphere. In such cases the following method may be used to extend the thresholds in fig. 35. *Fig. 36a* gives a block diagram of a conventional FM receiver, if the connection indicated by the dashed line is disregarded for the moment. The radio-frequency FM signal S of carrier amplitude A and modulation index α_1 (fig. 36*b*) is converted in the mixer M by the local oscillator LO into an intermediate-frequency signal S'. The relations in the vector diagram remain unchanged. The i.f. filter F_1 determines the bandwidth of the transmission channel, and hence at the same time the radius b_1 of the noise circle in fig. 36*b*. The FM detector D senses the frequency variations due to signal and noise, and the detected signal is limited in bandwidth by the lowpass filter F_2,

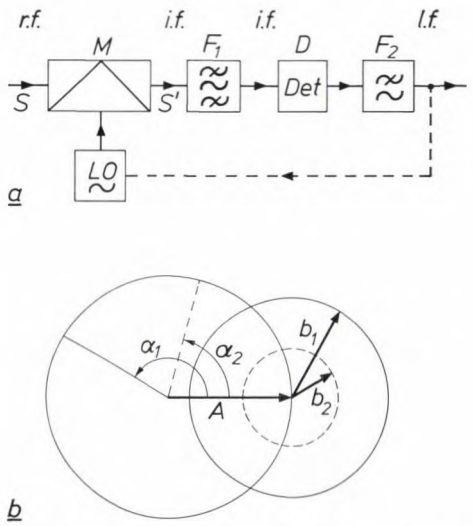


Fig. 36. *a*) Circuit for 'threshold extension'. M mixer, LO local oscillator, F_1 i.f. filter, D detector, F_2 low-frequency filter. In ordinary FM reception (no dashed line) the deviation of the i.f. signal S' is the same as that of the r.f. signal S. By means of the dashed connection the signal of the local oscillator is made to follow the input signal S in frequency, thus reducing the deviation of S'. The response of F_1 can then be narrowed. The noise circle becomes smaller and the distance to the threshold becomes larger. *b*) Modulation index and noise circle before (α_1, b_1) and after (α_2, b_2) the threshold extension.

[6] See for example J. van Slooten, Philips tech. Rev. **22**, 352, 1960/61.

which removes all signals of frequency greater than f_{max} (the maximum frequency of the information signal).

Let the modulation index α_1 be so large that the bandwidth it requires causes b_1 to approach closely to A; the threshold is then reached. The connection indicated by the dashed line in fig. 36*a* represents feedback that introduces the desired threshold extension. *LO* now oscillates at a frequency that depends on the instantaneous value of the signal from F_2, in such a way that the instantaneous frequency of S is followed. Consequently the frequency swing of S' becomes smaller, so that in fig. 36*b* the signal is now represented by, say, α_2. This does not cause any change in the noise (at least if S was not filtered before M). However, the filter F_1 can now be made so narrow that S' is only just accepted. As a result the noise is reduced, say to the circle of radius b_2 in fig. 36*b*. Thus, while preserving the signal-to-noise ratio, the threshold has been effectively extended. In this way it is possible in practice to shift the threshold in fig. 35 by 6 to 10 dB towards the left. There are limits, however, to the process. Narrowing of F_1 implies frequency narrowing and hence a delay in the control loop. In the long run *LO* follows the signal S with a delay such that the frequency deviation of S' remains too large for F_1.

Applications; magnetic recording of video signals

In the foregoing the emphasis has been placed on the possibility of obtaining a better signal-to-noise ratio with FM than with AM. This is of great importance in many forms of telecommunication. The insensitivity to amplitude variations is another great advantage, for example in microwave links and mobile installations. We shall not consider the use of FM in telecommunication further now, but will return to it at some length in the final section.

Another example of an application is to be found in frequency-analog measurement and control systems where quantities in a physical process to be telemetered or used in a remote-control system are converted into the frequency of an electrical signal, to protect the information from transmission interference [7].

In this section we shall confine ourselves to an example of the use of modulation for information storage: the magnetic recording of video signals. The reason for using FM here is its insensitivity to amplitude variations; there is no question of any gain in signal-to-noise ratio.

An audio signal on a gramophone disc or audio magnetic tape is recorded directly in the baseband. This cannot be done with a video signal, because the ratio of the highest to the lowest frequency is too great. The ratio that can be handled by a magnetic tape is of the order of 1000. This is given by the minimum and maximum wavelengths of the magnetic variations on the tape, which in their turn are determined respectively by the gap length of the head and the size of the head itself (*fig. 37*). At a maximum frequency of 5 MHz — the upper limit of the video signal (see *fig. 38*) — the lowest frequency would therefore be 5 kHz, and this is not low enough.

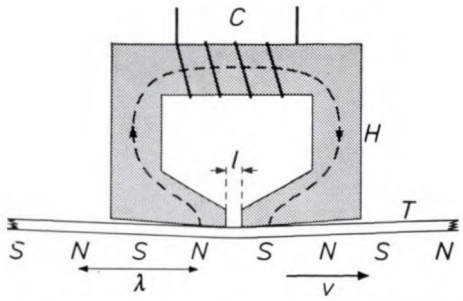


Fig. 37. Magnetic head H and tape T of a tape recorder, schematic. The interaction between the coil C and the tape magnetization takes place via the lines of force through the head. The interaction becomes very poor when the wavelength λ on the tape becomes smaller than the gap length l or greater than the head itself.

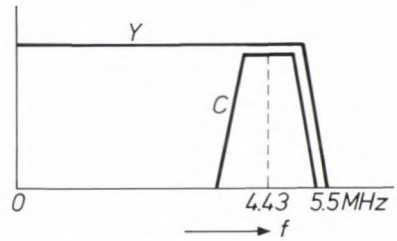


Fig. 38. Video baseband spectrum. The video signal consists of the luminance signal Y and the chrominance signal C (colour signal on a subcarrier at 4.43 MHz) [8].

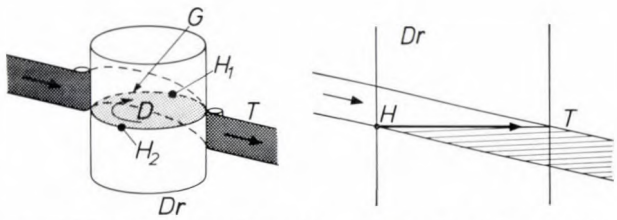


Fig. 39. Video tape recorder with two moving heads in the VCR, schematic. *Left:* the tape T follows a helical line around the drum *Dr*. The heads H_1 and H_2 rotate on the disc D along the gap G. *Right:* drum and tape developed into a plane. Each portion of track from one edge of the tape to the other corresponds to one television field period (1/50 s). The pitch of the tracks is 0.187 mm, the tape speed 14.3 cm/s, the head-over-tape speed 8.1 m/s.

Modulation is the remedy here: by shifting the video-frequency band towards higher frequencies, we reduce the frequency ratio while preserving the bandwidth. In particular, FM is the best method here because the signal amplitude at the highest frequencies (the shortest wavelengths) depends closely on the distance between head and tape and on the tape quality, and consequently can show considerable unintended variations.

The move to higher frequencies has to be paid for by a higher tape speed relative to the head. In tape recorders for use in the home (VCR, video cassette recorders) the video baseband cut-off is at 3 MHz, and a tape speed of 8.1 m/s relative to the head is used; this is

[7] D. Gossel, Philips tech. Rev. **34**, 288, 1974.

[8] Colour-television systems are discussed in F. W. de Vrijer, Philips tech. Rev. **27**, 33, 1966.

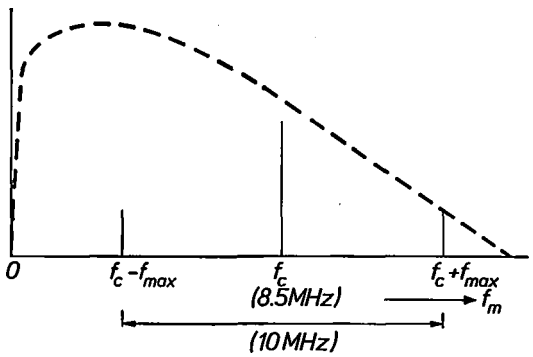


Fig. 40. The spectrum available with magnetic tape (schematically indicated by dashed curve) — which may differ slightly from tape to tape — is so limited that it can barely contain at the same time two sidebands of the first order, $f_c \pm f_{max}$ (f_{max}, the width of the baseband, is 5 MHz in a professional video tape recorder). In this diagram the choice of 8.5 MHz for the 'carrier frequency' f_c is somewhat arbitrary.

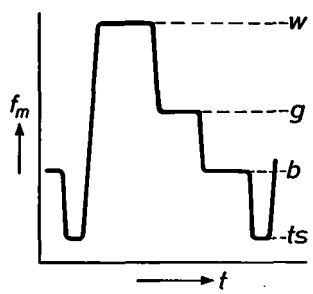


Fig. 41. Instantaneous frequency f_m of the modulated signal as a function of time during a picture line period, for the case where a white, a grey and a black area are side by side in the picture. The table below gives the values of f_m for 'white' (*w*), 'black' (*b*) and 'top sync' (*ts*), the passed bandwidth f_{max} of the video signal, the maximum peak-to-peak frequency deviation $2\Delta f$, the modulation index $\alpha = \Delta f / f_{max}$ and the head-over-tape speed h in broadcast equipment and in VCR.

	Broadcast	VCR
f_m white	9.30 MHz	4.4 MHz
black	7.80 MHz	3.5 MHz
top sync	7.16 MHz	3.0 MHz
f_{max}	5.0 MHz	3.0 MHz
$2\Delta f$	2.14 MHz	1.4 MHz
α	0.21	0.23
v	41.1 m/s	8.1 m/s

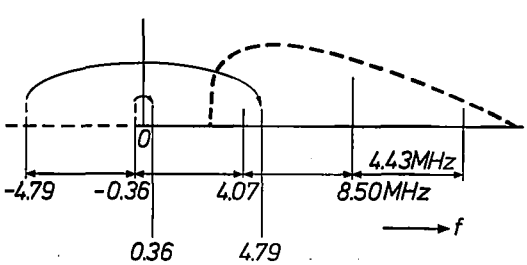


Fig. 42. FM spectrum at a 'carrier frequency' of 8.50 MHz and a baseband signal of 4.43 MHz, the frequency of the colour subcarrier. Since in practice negative frequencies are equivalent to positive ones, lower sidebands with negative frequencies are found 'folded back' into the spectrum of positive frequencies. In the present case the third lower sideband comes into the band used (shown dashed) and would cause interference there if α were not so small.

achieved by means of two moving heads (*fig. 39*). In professional equipment used for broadcasting the baseband goes up to 5 MHz; the head/tape speed of 41.1 m/s is obtained with four moving heads.

The available bandwidth is unfortunately so small that it is barely able to contain the $2f_{max}$ part (10 MHz in professional equipment) of the $2f_{max} + 2\Delta f$ bandwidth required according to Carson (*fig. 40*). Thus little remains for $2\Delta f$, the maximum peak-to-peak frequency deviation; see *fig. 41*. The modulation index α is therefore much smaller than 1, so that there is no signal-to-noise improvement.

A very small α in the situation illustrated in fig. 40 is also required for other reasons. Since the 'carrier frequency' f_c is not much greater than half the bandwidth — unlike the case in FM broadcasting, for example — 'folded lower sidebands of higher order' enter the frequency band used, and would give rise to interference if α was not so small. In this respect the subcarrier of the colour signal (fig. 38), which is always present at a fairly large amplitude, is especially troublesome; see *fig. 42*. A further danger comes from the nonlinearity of the recording process. This gives rise to higher harmonics of the carrier in the read-out signal (at frequencies $2f_c$, $3f_c$, ...), which are also frequency-modulated, and whose lower sidebands may enter the band in use. For this reason again α must be small.

In fact the process of magnetically recording video signals is very difficult to analyse. In the first place the 'carrier frequency' does not have such a distinct meaning as it has in FM broadcasting, because the video signal, and hence the instantaneous frequency f_m, have no obvious mean: it is lower in the darker areas than in the light areas of the picture. Here f_c can best be regarded as the mean of f_m over suitably chosen short periods.

In the second place the situation is complicated by the fact that the available frequency band (fig. 40) is very limited, is not flat and is different for every magnetic tape. In particular, deviations from the normal FM situation arise because the upper sideband is attenuated more than the lower sideband. The magnitude of this effect depends both on the baseband frequency (which may have values from 0 to f_{max}) and on the 'carrier frequency' f_c. These deviations, however, are partly overcome by a VSB-type effect arising as a result of the sloping band edge (see figs. 20 and 21). In any case, these effects are least pronounced at the low frequencies of the baseband signal, the frequencies to which the eye is most sensitive.

In the video cassette player, as stated earlier, the baseband signal is limited to about 3 MHz. The colour subcarrier is therefore suppressed. The colour information in this case is carried in QAM on a separate carrier at about 560 kHz. The frequency band for this information is below the band used for the brightness signal. At these low frequencies the amplitude variations resulting from 'band flutter' are sufficiently small for this form of amplitude modulation.



RO 150 communications receiver for frequencies between 0.2 MHz and 30 MHz (wavelengths between 10 and 1500 m). The tuning is electronic; the coarse tuning is by stepping switches and the fine tuning is continuous over intervals of 100 kHz. The tuned frequency is displayed digitally. The receiver can detect various types of amplitude modulation: conventional AM, AM with suppressed carrier, upper-sideband and lower-sideband modulation. In addition a separate module permits the reception of AM with two independent sidebands (ISB), and of frequency-shift keying and facsimile.

II. Modulation of pulse trains

Sinusoidal waves are in a way the natural carriers of information, especially in radio communication, for several reasons, two of which will be mentioned here. In the first place the system of sinusoidal carriers allows the spectrum to be divided up into an orderly arrangement of frequency bands; by using a band filter one signal can always be selected from the innumerable number present in the radio spectrum. In the second place, the modulated sinusoidal carrier generally experiences very little distortion during transmission because it occupies only a relatively narrow frequency band. Broadband signals usually do suffer distortion, because in both cables and radio transmission Fourier components that differ considerably in frequency may also show considerable differences in attenuation or transit time.

In these respects rectangular pulse trains do not make suitable 'carriers' since they occupy an infinitely broad spectrum. For this reason they are not used as carriers for radio transmission. Where information modulating a pulse train is to be transmitted by radio, the pulse modulation is always followed by modulation of a sinusoidal carrier, leaving a free choice of frequency band. In cable transmission, again, a sine wave is often used as the pulse-train carrier.

Why, then, is pulse modulation used?

In telecommunication, and especially in telephony, there is a growing trend towards the large-scale transmission of signals both in *time-division multiplex* (TDM) and in *quantized and coded form*. Quantization and coding allow information to be transmitted virtually undistorted over large distances. Time-division multiplex, like frequency-division multiplex, allows many information signals to be transmitted simultaneously over a single path. Pulse trains modulated in amplitude are the foundation of these processes.

'Sampling' is the first operation to which an analog information signal is subjected when transmitted in TDM, whether or not quantized and coded. The resultant pulse train, modulated in amplitude by the signal, forms a series of 'samples' of the signal (*fig. 43*). The essential question concerning the number of samples per second that are necessary if no information is to be lost is answered by the Nyquist theorem. The sampling of analog signals, incidentally, is a general process that is widely applied in signal-handling equipment, for example in CTDs ('charge-transfer devices').

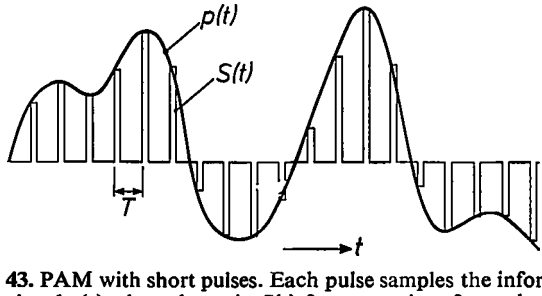


Fig. 43. PAM with short pulses. Each pulse samples the information signal $p(t)$; the pulse train $S(t)$ forms a series of samples.

Quantization and coding are the subjects of part III. The other subjects just mentioned are based on pulse-*amplitude* modulation, and will be dealt with shortly. We shall then look at pulse-*frequency* and pulse-*duration* modulation with reference to the 'VLP' record. First of all we shall look briefly at a number of basic concepts.

Some basic concepts

An unmodulated train of rectangular pulses is characterized by the amplitude A, the pulse frequency f_c, the phase ϕ and the duration τ of the pulses (*fig. 44*). Modulation of these quantities gives pulse-amplitude modulation (PAM), pulse-frequency modulation (PFM), pulse-phase or pulse-position modulation (PPM) and pulse-duration modulation (PDM). As in the case of sine waves, variations in phase are accompanied by variations in frequency, and vice versa. Here again we speak of frequency modulation where the frequency variations are proportional to the information signal, and of phase modulation where this is the case with the phase variations (fig. 44*c* and *d*).

As noted earlier, an unmodulated pulse train (fig. 44*a*) occupies an infinitely broad spectrum. Its Fourier expansion is:

$$P(t) = \beta A + \frac{2}{\pi} A \sum_{n=1}^{\infty} \frac{1}{n} \sin n\pi\beta \cos n(\omega_c t + \phi). \qquad (18)$$

Here ω_c is equal to $2\pi f_c = 2\pi/T$ (T being the pulse repetition period), and β, the 'pulse duty factor' is equal to τ/T. The phase ϕ determines the 'position' of the pulses, e.g. with respect to the time $t = 0$. (For $\phi = 0$, $t = 0$ at the centre of a pulse.)

The character of the spectrum is determined by the value of β; for example, if $\beta = \frac{1}{2}$, all even harmonics are zero. As an example *fig. 45* gives the spectrum for $\beta = 0.4$. This illustrates the important fact that the

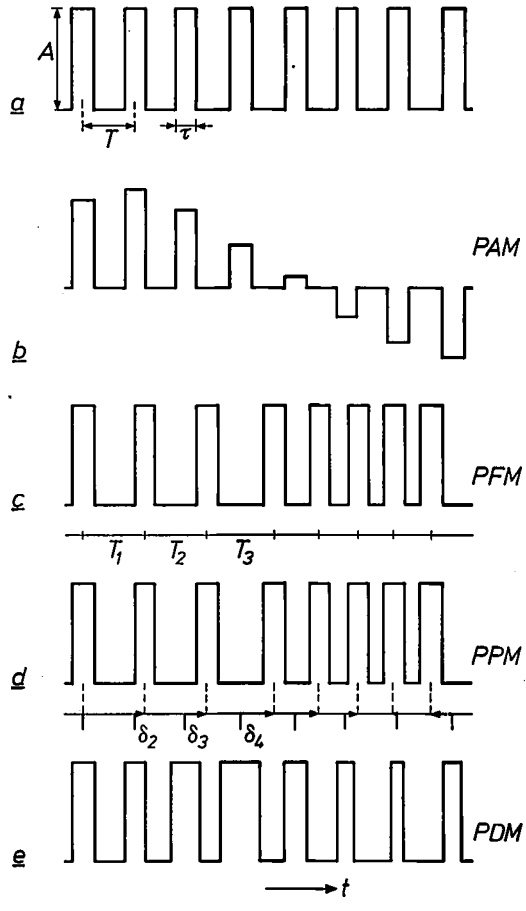


Fig. 44. *a*) Train of rectangular pulses. A amplitude, T period, τ pulse duration. The ratio $\tau/T = \beta$ is known as the 'pulse duty factor'. The pulse frequency f_c is equal to $1/T$. The 'pulse position' δ, i.e. the distance from the centre of a pulse to the last preceding moment nT, is equal to $-\phi/\omega_c$, where ϕ is the phase (see eq. 18). In PAM (*b*), PFM (*c*), PPM (*d*) and PDM (*e*), A, T, δ and τ are functions of time. The identical pulse trains *c* and *d* are said to be PF- or PP-modulated, depending on whether the variations in f_c, and hence in $1/T$ or those in ϕ (and hence those in δ) are proportional to the information signal.

amplitude of the nth Fourier component tends to zero when n increases without limit. This means that the pulse shape in fig. 44a can be reasonably approximated in a frequency band of finite width; the broader the band the better the approximation. In particular, the *rise time* τ_r — the time in which the signal value increases from 10% to 90% of the peak pulse amplitude, and which is zero in the ideal case — is shorter

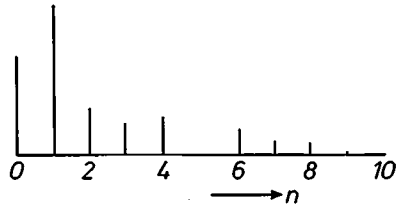


Fig. 45. Spectrum of an unmodulated pulse train for which the pulse duty factor is 0.4.

the broader the available band B. The relation between τ_r and B is

$$\tau_r \approx 1/2B, \qquad (19)$$

as demonstrated roughly in *fig. 46*.

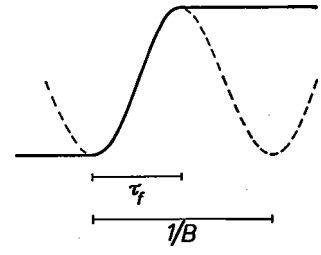


Fig. 46. The Fourier component with the highest frequency of a step signal determines both the rise time τ_r of the signal and the bandwidth B of the spectrum. If we assume that this component (dashed) must fit the step (solid line) as shown here, we see that τ_r must be about half the period $1/B$ of the component.

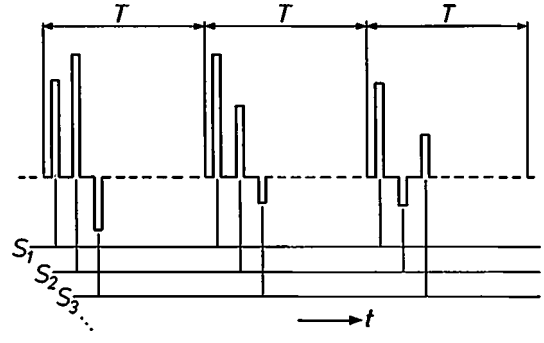


Fig. 47. Time-division multiplex (TDM). The successive time intervals ('frames') of duration T each contain one pulse of each signal S_1, S_2, S_3, ...

Pulse-amplitude modulation (PAM)

Time-division multiplex (TDM)

The idea of time-division multiplex (TDM) is that other signals can be transmitted in the intervals between the sampling pulses of a signal (see fig. 43). The time is thus divided into equal intervals of length T, each of which contains one pulse of each signal (*fig. 47*). On reception the pulses have to be sorted again, which requires synchronization of the receiver with the transmitter. As we shall presently see, the number of signals with a given baseband width b that can be transmitted in this way over a transmission channel of bandwidth B is equal to $B/2b$, that is to say just as large as in FDM with DSB. (In FDM with SSB the number is twice as large.) Depending on the application, however, time-division multiplex with PAM can have a great advantage over frequency-division multiplex with DSB or

SSB. This is because frequency division requires the use of *filters*, which often take up a lot of space and need much attention (alignment, stabilization), whereas the electronic switches required for time division are usually smaller and simpler. This advantage is particularly apparent in telephone exchanges. For this reason TDM is used in small telephone systems, with the pulse trains directly sent out after sampling, i.e. uncoded. The application of TDM to coded signals will be dealt with in part III.

The maximum number of signals that can be transmitted is equal to the maximum length of the time interval T (fig. 47) divided by the minimum width of the pulses. This minimum width is twice the rise time and is therefore equal to $1/B$, from (19). To determine the maximum length of T, we must look at the spectrum of the modulated wave $S(t)$.

The modulated pulse train $S(t)$ in fig. 43 may be regarded as the product of $p(t)$ and an unmodulated pulse train $P(t)$ (eq. 18):

$$S(t) = p(t)P(t) = \\ = \beta A p(t) + \frac{2}{\pi} A \sum_{n=1}^{\infty} p(t) \frac{1}{n} \sin n\pi\beta \cos n\omega_c t.$$

Each harmonic component is thus modulated in amplitude by the information signal $p(t)$ and so has two sidebands (*fig. 48*). This means that each harmonic occupies a total bandwidth of $2b$. It is now clear that the information signal can be recovered from $S(t)$ with the aid of a lowpass or bandpass filter, *provided* the components do not overlap. This condition gives the maximum value for T, or the minimum value for the pulse frequency f_c ($= 1/T$), which must be at least $2b$. The maximum number of signals that can be transmitted in time-division multiplex, $T_{max}/(1/B)$, is thus indeed $(1/2b)/(1/B) = B/2b$. We should note in passing that the first step of the filtering operation by means of which the information signal is reconstructed from the pulse train obtained after reception and sorting simply consists in holding the sample value until the next one arrives ('sample and hold', *fig. 49*).

Nyquist's theorem

It could perhaps appear possible that there might be a clever way of recovering the information signal even when the bands in fig. 48 do overlap. In fact, this cannot be done, since the requirement '$f_c \geqslant 2b$' is a special form of *Nyquist's theorem*, which states that to define a signal of bandwidth b at least $2b$ independent quantities per second are necessary. This can be explained as follows. Let us consider a signal $p(t)$ over the time interval from t_0 to $t_0 + t_1$. We expand this as a Fourier series with components at the frequencies $1/t_1, 2/t_1, \ldots, n/t_1$. The highest frequency in this series is also the bandwidth: $n/t_1 = b$. Each component is fixed by two parameters, e.g. the amplitude of the sine phase and that of the cosine phase. The total of numbers characterizing the function is thus $2n = 2bt_1$, and the total required per second is thus $2b$.

If we nevertheless sample at a frequency f_c smaller than $2b$ ('sub-Nyquist sampling'), then the components at frequencies between $f_c - b$ and b from the lower sideband of the fundamental will also be passed by the lowpass filter and cause interference ('aliasing'). A well known instance of this situation is found in the occurrence of Moiré patterns in superimposed patterns of parallel lines (or adjacent sets of railings), with only slightly different numbers (k_1 and k_2) of lines or railings per metre. The lower pattern (k_2) is effectively sampled in position instead of in time by the spacings in the upper pattern (k_1). If k_1 is much greater than k_2, a good picture of the lower pattern is obtained. If this is not the case, however, and k_2 is on the contrary almost equal to k_1, then the viewer sees mainly Moiré fringes at a spatial frequency of $k_1 - k_2$ per metre. In television pictures Moiré fringes can occur between a pattern of horizontal stripes in the subject and the lines that scan the picture. The line scan here constitutes sampling in the vertical direction.

In television there is no question of horizontal sampling of a video signal in the ordinary sense. This is however the case where the video signal is to be transmitted in digitally coded form (see part III). To keep the sampling rate f_c as low as possible in this case, sub-Nyquist sampling is applied where the subjective effects remain below a limit considered to be permissible. The effects (in a stationary picture) consist of stripes with a periodic spacing that corresponds to the frequency $f_s = f_c - f_0$ of the interference. Here f_0 is a desired frequency in the video signal. With a suitable choice of f_c, for example $(n + \frac{1}{2})f_L$, where f_L is the line frequency and n an integer, the subjective effect can be

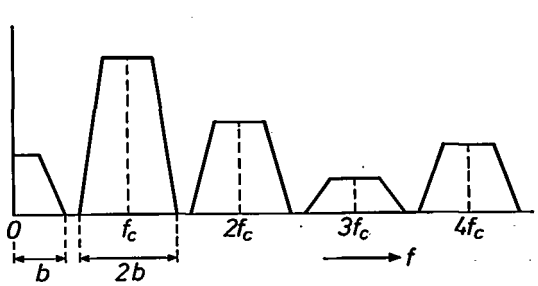


Fig. 48. Spectrum of amplitude-modulated pulse train, schematic. Each Fourier component of the unmodulated pulse train (see for example fig. 45) gets two sidebands.

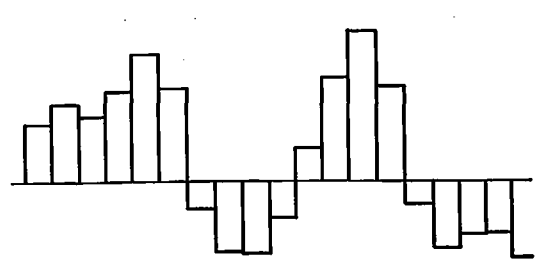


Fig. 49. To recover the information signal from a pulse train $S(t)$ (see fig. 43) it is common practice to hold each sample value until the next sample value arrives ('sample and hold'). Most r.f. components of $S(t)$ have then been filtered out.

considerably reduced. Since f_0 is a multiple of the picture frequency, and is also usually close to a multiple of f_L, the frequency f has approximately the form $(m + \frac{1}{2})f_L$ (m an integer), so that the interference is always approximately in antiphase with that in the preceding line. Since the number of television lines is odd, the interference is also in antiphase in successive pictures. As a result the interference is hardly visible, partly because of the integration time of the eye and partly because of the break-up of the striped structures.

Another example of an application where the sampling of analog signals is essential is to be found in charge-transfer devices (CTDs), an invention that has been used more and more widely in recent years in many kinds of signal processing [9]. In these devices successive samples of the signal are stored in the form of a charge in a series of cells, and are passed along at a rate fixed by an external clock (*fig. 50*). The samples are detected at the output and reconstituted into a signal, so that the device forms a delay line, whose delay can be varied by means of the clock frequency. Variable delay lines of this type are used, for instance, for correcting time errors that arise in the playback of television signals stored on magnetic tape, and for generating artificial reverberation in audio installations. Charge-transfer devices are also used in 'transversal filters' (*fig. 51*). Here the output signal at any instant consists of the sum of the sampled values in the memory device, after each of them has been multiplied by a particular factor. If the factor is put equal to 1, for example, then the output signal is always proportional to the mean input signal over a period corresponding to the length of the device, so that we have a lowpass filter. With a suitable choice of factors any possible filter characteristic can be produced.

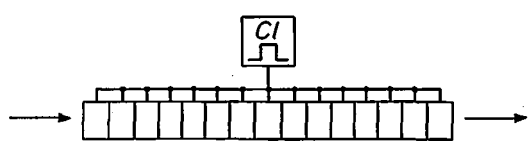


Fig. 50. Charge-transfer device (CTD). Incoming samples are stored in the form of a charge in the cells and are shifted along one step with every pulse from the clock *Cl*.

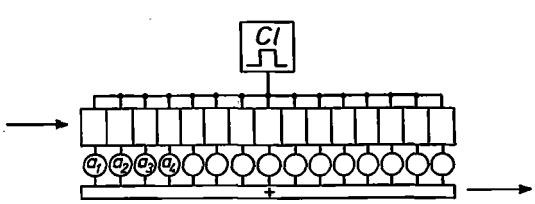


Fig. 51. Transversal filter. The sample values in the cells are all multiplied by an appropriate factor and added. If the factors are all set equal to 1, the circuit acts as a lowpass filter. Other choices lead to other filter characteristics.

If the samples in a CTD are to represent the signal faithfully, the sampling rate — in this case the clock frequency — must again be at least twice the bandwidth b of the signal. In the delay-line application the delay time is equal to the number of device elements N multiplied by the clock period. The delay time can thus be regulated by means of the clock frequency, but only up to a maximum of $N/2b$.

Pulse-frequency and pulse-duration modulation (PFM and PDM); the 'VLP' record

Pulse-frequency, pulse-position and pulse-duration modulation are all forms of what is sometimes known as pulse-time modulation. Features common to all of these are that there are only two levels for the signal value, and that the information is contained in the time-interval variations between the successive transitions between these levels. Because of this, in the ideal case of infinitely steep pulse edges, they are insensitive to interference and noise affecting the levels (provided that the interference signals do not exceed half the spacing between the levels); see *fig. 52*. In the case of

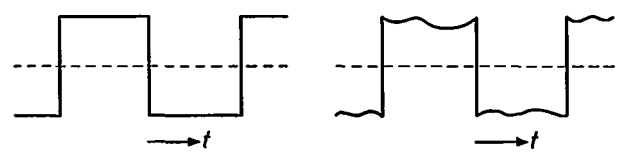


Fig. 52. In pulse-time modulation only the times of transition between two levels are important. The information is completely unaffected by additive noise, provided that the pulse edges are infinitely steep and the noise level is not so high that the signal occasionally exceeds the discrimination level (dashed line).

non-vertical pulse edges, the time error Δt caused by an additive interference a in the signal value (see *fig. 53*) is given by:

$$\Delta t/\tau_t \approx a/A.$$

From (19) τ_t is equal to $1/2B$, and therefore:

$$B\Delta t \approx a/2A. \qquad (20)$$

Equation (20) expresses the important feature of time modulation that we have already discussed in some detail under the time modulation (FM) of sine waves, i.e. that the effect of additive interference can be reduced by increasing the bandwidth. This applies when a is independent of B, as in the case for example of a single interference. In the case of white noise the noise power is proportional to B, and a is therefore proportional to $\sqrt{B}$, so that there is still a reduction in the effect of the noise to be obtained by increasing the bandwidth.

The 'VLP' record [10] is an interesting example of an application using a combination of PFM and PDM, and in the rest of this section we shall concentrate on this application. Played on the appropriate equipment, the 'VLP' record reproduces a television programme of about half an hour, complete with colour picture and stereo sound. All the information in the programme is contained on the record, in a 25-km long spiral track with a pitch of 2 μm between circles of diameter 15 and 30 cm. The track consists of tiny pits 0.8 μm wide and 0.16 μm deep. An optical scanning system and a photocell produce trains of electrical pulses from the pits. With a pulse frequency of 6.3 MHz and a disc speed of 25 rev/s (1500 rev/min) the pit period on the track diameter of 30 cm must be about 3.8 μm (and hence 1.9 μm on the track diameter of 15 cm). The brightness information is stored by PFM, i.e. in the variations in the spacings between the centres of the pits. The brightness at a point of the television picture corresponds to a particular value of the instantaneous frequency; black corresponds to 6 MHz and white to 7 MHz. The instantaneous frequency of 5.6 MHz is reserved for the line-synchronization signal ('top sync'). The colour and sound information is stored by PDM, i.e. in variations in the length of the pits [11].

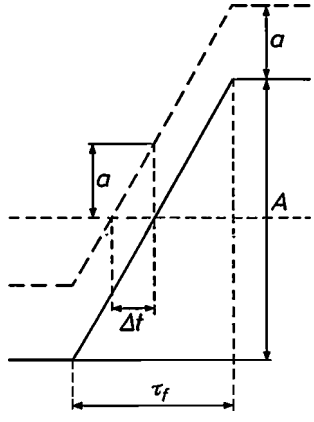


Fig. 53. Pulse edge of finite steepness, without interference (solid line) and with additive interference a (dashed line). The time error Δt is a/A times the rise time τ_f, and therefore the smaller the steeper the edge.

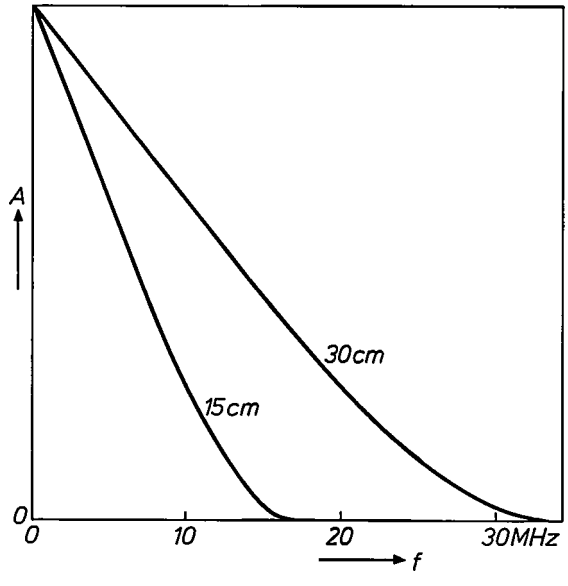


Fig. 54. The amplitude of the read-out signal during playback of a 'VLP' record, as a function of the frequency of the recorded pulse train. At high frequencies the amplitude shows a marked decrease because the scanning spot is no longer able to discriminate between the pits in the disc. Owing to the difference in linear velocity the critical distance between pits on the track at a diameter of 15 cm corresponds to a lower frequency than on the track at a diameter of 30 cm.

A two-level system with time modulation is the obvious form of information storage here, because of the strongly nonlinear photographic process used for tracing the track and the unavoidable variation in level during read-out. The possibility of reducing the effects of interference by widening the bandwidth is not really used. On the contrary, a band of only about 10 MHz is available. This is determined by the size of the light spot of the scanning system; it is not able to discriminate between pits that are too short or too close together, so that above a particular frequency — which depends on the track diameter — the read-out signal disappears (*fig. 54*). To accommodate all the information in the available band the brightness signal is limited in baseband width to 3 MHz. *Fig. 55* gives a diagram of the processes applied to the luminance, chrominance and audio signals before they are recorded on the disc. What we are concerned with here is only the last step, the PFM and the PDM of the track of pits.

Pulse-frequency modulation amounts to variation of the quantity ϕ in eq. (18); this will also have the effect of varying the period of the pulses and hence the 'instantaneous value' of the pulse frequency (as with sine waves, variation of the parameter ω_c in (18) would be pointless). We thus have frequency modulation of each Fourier component, so that the total spectrum consists of a superposition of FM spectra in the way illustrated in fig. 27, with the centre frequencies f_c, $2f_c$, $3f_c$, ...

[9] The 'bucket-brigade delay line' was the first of the storage devices now grouped under the heading of CTD. See F. L. J. Sangster, Philips tech. Rev. 31, 97 and 266, 1970, and L. J. M. Esser, Electronics Letters 8, 620, 1972.

[10] See the four articles on the video long-play ('VLP') system in Philips tech. Rev. 33, 178-193, 1973.

[11] These are only numerical examples and are far from definitive. It has also been found that it is possible to record the colour picture signal, including the colour subcarrier modulated by the colour signal, as a single entity in PFM on the disc. (The sound is still applied by PDM.) This requires a larger bandwidth. It has been decided to do this for the American colour-television signal (NTSC system), but no decision has been reached on this question in Europe.

The bandwidth B_n of the nth spectrum is given by the rule of thumb (16):

$$B_n = 2(n\Delta f + f_{max}),$$

where f_{max} is the bandwidth of the baseband signal and $n\Delta f$ is the frequency deviation of the relevant component. In the case of the 'VLP' disc, f_{max} is equal to 3 MHz. The value chosen for the frequency deviation Δf of the fundamental is 0.7 MHz, so that B_1 is about 7.4 MHz and B_2 about 8.8 MHz. Overlapping would thus occur between the spectrum of the fundamental (6.3 MHz) and that of the second harmonic (12.6 MHz)

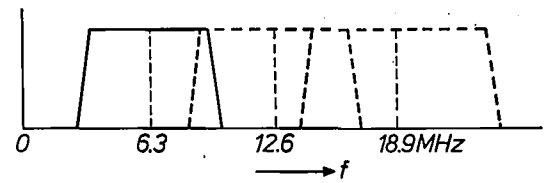


Fig. 56. Band spectrum of a frequency-modulated pulse train (pulse frequency of unmodulated signal 6.3 MHz, maximum frequency of baseband spectrum 3 MHz, frequency deviation of the fundamental 0.7 MHz). Each harmonic produces an FM band. The second harmonic does not appear on the 'VLP' disc, owing to the choice of $\frac{1}{2}$ for the pulse duty factor. The bands of the third and higher harmonics, if still present in the read-out signal, are filtered out.

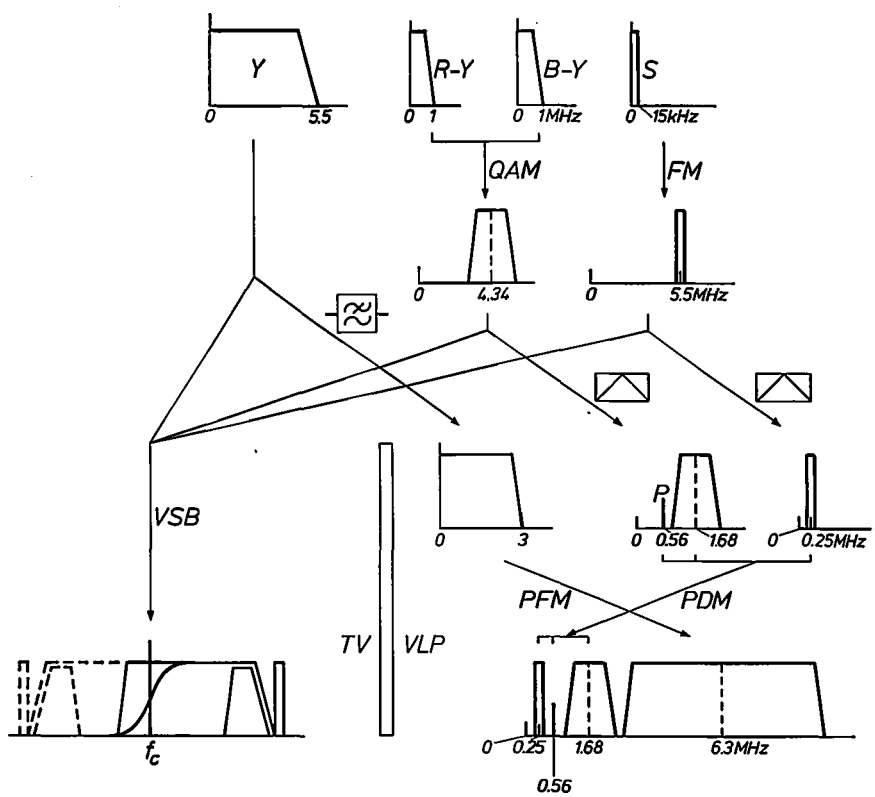


Fig. 55. Operations on the luminance signal Y, the chrominance signals $R-Y$ and $B-Y$ and the sound signal S before they are transmitted or recorded on a 'VLP' disc. Chrominance and sound modulate subcarriers by QAM and FM, respectively. In a television transmission these signals, together with the luminance signal, modulate a carrier of say 197 MHz in VSB. For recording on the 'VLP' disc the luminance signal is limited in bandwidth, and the chrominance and sound carriers are shifted in frequency by mixing. A pilot signal (P) is added for synchronous detection of the QAM chrominance signal. Finally a pulse train, which has a pulse frequency of 6.3 MHz unmodulated, is modulated in frequency by the luminance signal and in pulse duration by chrominance, sound and the pilot signal. The way in which this is done is explained in fig. 57.

(see *fig. 56*), if it were not for the fact that *the second harmonic is missing, since the value $\frac{1}{2}$ has been chosen for the (mean) duty factor β*; the spacing between the pits is thus (on average) just as long as the pits themselves. The spectra of the third and higher harmonics are completely suppressed. We are only concerned, therefore, with the spectrum of the fundamental. It should be noted that, as distinct from the usual situation in FM radio, the ratio of frequency deviation (Δf) to baseband width (f_{max}), i.e. the modulation index, is considerably *smaller* than 1, i.e. 0.7/3 or 0.23. This again makes it clear that there is no question of any noise suppression by band broadening.

As stated earlier, the colour and sound information is stored by PDM, that is by modulation of the quantity β in (18). As can be seen in eq. (18), and as also

appears directly from fig. 44*e*, variation of β implies variation of the 'd.c. terms' of the modulated signal. The spectrum of the modulating signal thus enters the spectrum of the modulated signal unchanged. On the 'VLP' disc the colour and audio bands, transposed by mixing to relatively low frequencies, fit exactly into the hole left by the clipped brightness band in the spectrum (see fig. 55).

Finally, *fig. 57* illustrates schematically how the PF- and PD-modulated signal recorded on the disc is made.

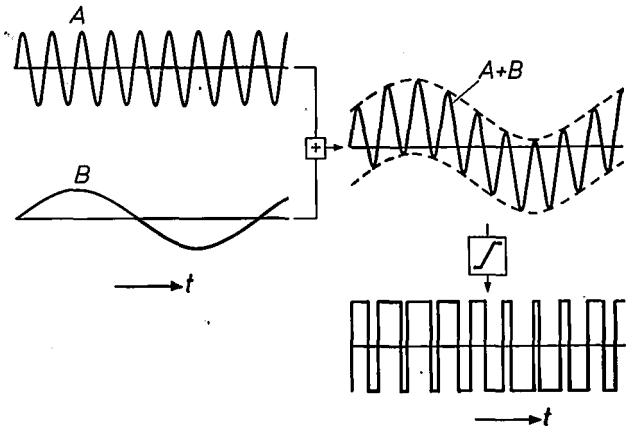


Fig. 57. Method of producing a pulse train of frequency f_c that is modulated in frequency by a signal Y and in pulse duration by a signal B. (In the 'VLP' disc, f_c is equal to 6.3 MHz, Y is the luminance signal and B the combination of modulated chrominance subcarrier, modulated sound subcarrier and pilot signal; see fig. 55.) In the figure A stands for a sine wave of frequency f_c modulated in frequency by Y. Addition of A and B and limiting of the sum obtained yields a pulse train that contains the frequency variations of A and with pulse durations approximately proportional to B. The frequency of B must be appreciably lower than the frequency of A, and the amplitude of B must be appreciably smaller than the amplitude of A.

Since β occurs not only in the d.c. terms but also in all the other components, variation of β has other effects as well. It will be shown that in this respect also the value of $\frac{1}{2}$ is a particularly good choice for β. For the 'VLP' disc only the constant and the fundamental in (18) are relevant, and we shall confine ourselves to these. Let us put

$$\beta = \tfrac{1}{2} + \{\gamma + y(t)\}/\pi, \qquad (21)$$

where γ/π is a constant deviation of β from the ideal value $\frac{1}{2}$, and $y(t)$ is proportional to the modulating signal. We assume that $\{\gamma + y(t)\}/\pi$ is much smaller than 1. Substituting (21) in (18) then gives:

$$\begin{aligned} P(t) &= (\tfrac{1}{2} + \gamma/\pi + y/\pi)A + (2/\pi)A\cos(\gamma + y)\cos\omega_c t \approx \\ &\approx (\tfrac{1}{2} + \gamma/\pi + y/\pi)A + (2/\pi)A\{1 - \tfrac{1}{2}(\gamma + y)^2\}\cos\omega_c t = \\ &= (\tfrac{1}{2} + \gamma/\pi + y/\pi)A + (2/\pi)A(1 - \tfrac{1}{2}\gamma^2 - \gamma y - \tfrac{1}{2}y^2)\cos\omega_c t. \end{aligned}$$

If y is a sinusoidal signal of frequency f_p, we find in the spectrum, in addition to components at f_p and f_c, interfering components at $f_c \pm f_p$ and $f_c \pm 2f_p$, deriving from the terms $\gamma y \cos\omega_c t$ and $y^2 \cos\omega_c t$ respectively. Of these at least the first is missing if γ is zero and β thus has on average the ideal value of $\frac{1}{2}$; the components at $f_c \pm 2f_p$ are only of the second order in y.

It requires considerable mastery of the art of disc-manufacture to ensure that β remains on average sufficiently close to $\frac{1}{2}$. The most undesirable interfering component is the colour subcarrier at 1.68 MHz (see fig. 55). The first-order interference from this acts as a sideband at frequency $f_m \pm 1.68$ MHz of the FM brightness signal at the instantaneous frequency f_m, and is found to give a striped pattern with a spatial frequency corresponding to 1.68 MHz. The second-order interference ($f_m \pm 3.36$ MHz) falls outside the bandwidth of the PFM signal and is filtered out. The audio and pilot signals are weaker and therefore less of a nuisance.

Finally we should note that, in principle, β and ϕ in (18) are independent quantities, so that, again in principle, PDM and PFM can be independent of one another and therefore it should not be necessary to separate them in frequency. Frequency separation is necessary, however, because the nonlinear signal processing and the bandwidth restriction degrade the isolation and introduce crosstalk.

* *
*



A Mariphone transceiver aboard a yacht. The Mariphone service of the Netherlands Post Office allows telephone calls to be made from boats to any subscribers to the public telephone system and is the counterpart of the Mobilophone service available to road users. Sea-going vessels can also use Mariphone in coastal waters. The system has its own frequency channels (in the 156-162 MHz band); the modulation is FM.

III. Quantization and coding of analog signals

An audio signal from a microphone or a video signal from a camera will become degraded by noise or other interfering signals when it is transmitted over a particular channel, either in baseband or on a carrier. On reception it is no longer possible to separate the original signal from the interference imposed on it. This is because the signal is an 'analog' signal: within certain limits it can assume any value. When a signal is *quantized*, however — that is to say given a range of discrete values — the receiver will always be able to discriminate between the signal values in spite of the added interference, provided that the interfering signals are not too large compared with the differences in the signal levels. This opens up the possibility of almost completely eliminating the effects of noise and interference. The advantage when signals have to be transmitted over great distances is particularly apparent, because as the distance increases an analog signal inevitably accumulates more and more noise and interference, so that the signal is finally lost in the noise, whereas a quantized signal can be regenerated at repeater stations.

To take advantage of this technique for analog signals from a microphone or camera, the signals have to be transformed into quantized signals that carry the same information. Usually this is done in such a way as to result in a digital signal; the process is then called 'digital coding'. In particular, in 'binary coding' the result is a 'bit train', a train of pulses that can only have two values ('0' or '1').

Digital coding is now used on a large scale and to a growing extent in telephony. Apart from the advantage of a signal quality that is virtually independent of distance, it also provides the possibility of transmitting signals in time-division multiplex (TDM), with the benefits we have already seen. The system of digital coding in combination with TDM was first used on overcrowded lines between exchanges at short and medium distances, but it is now increasingly being used for long-distance communications. Owing to the high investment costs, the public telephone network is only gradually being digitized and it will probably take twenty or thirty years or more before it goes completely digital. In military communication, where digital coding is also widely used, the advantages of secrecy and the possibility of changing the code carry particular weight. For television digital coding is still in the experimental stage. In Britain there are experimental digital-coded television links between Southampton, Portsmouth and Guildford.

In this part of the article we shall take a closer look at a number of coding methods.

Pulse-code modulation (PCM)

A straightforward development of the idea outlined above is to be found in 'pulse-code modulation', a name that is not really correct since what is involved is coding and not modulation. The coding takes place in three stages in each sampling interval; see *fig. 58*. First of all the analog signal is sampled; the sampling rate must of course be sufficiently high. Next, the sampled value is quantized, i.e. rounded off to the nearest level in a preselected scheme of levels. If the levels are equidistant, we have linear PCM. The levels are thought of as being numbered, and the number obtained is finally coded in a pulse train. Notation of the numbers in the binary system results in binary coding. If for example we reserve 8 bits per number, we can operate with 256 levels. We then say that we have 'words' of 8 bits. Out of the bit train received the receiver must of course 'know' which of each 8 bits together form a word; this requires 'word synchronization'. This does not have to be provided for each word separately but only now and then, for example at regular intervals after a certain number of words. Examples of sampling and bit rates in use are given in *Table I*.

The sequence of level values obtained after decoding in the receiver is an approximation to the sequence of sampling values. After having passed through a low-pass filter the signal is an approximation to the original signal. The error arising from the difference between the level values and the sampling values is known as 'quantization noise'.

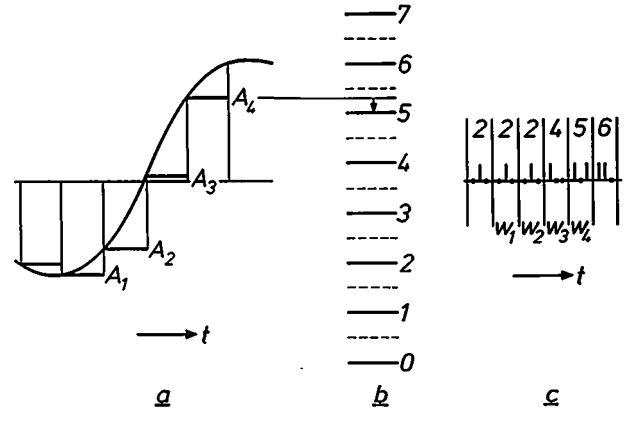


Fig. 58. Coding of an analog signal in PCM. *a*) The analog signal is sampled. *b*) The samples ..., A_1, A_2, A_3, A_4, ... are each rounded off to the nearest level of a given scheme of levels (quantization). *c*) The sequence of level numbers obtained is converted into a sequence of 'words' ..., w_1, w_2, w_3, w_4 ... which together form a pulse train. In the figure the scheme consists of 8 levels, which can be coded in binary form into words of three bits. The sample A_4, for example, yields level 5 and the word 101 (w_4).

The fineness of the scheme of levels chosen is a compromise. A coarse scheme gives a great deal of quantization noise. A fine scheme requires a large wordlength, and hence a high bit rate and a broadband transmission channel. A good compromise between quantization noise and wordlength can be reached with a nonlinear level scheme. For speech signals an optimum, nonlinear level scheme has been internationally standardized.

selves, we can either divide up the region much more finely with the same number of levels, and hence reduce the quantization noise, or divide it up just as finely with far fewer levels, which implies a reduction in the bit rate required. The advantage gained here is paid for with a 'rounding off' of large transitions between the samples ('slope overload'), since a large change in signal level can only be covered by several steps; see *fig. 60*.

Table I. Examples of sampling rates f_c and bit rates f_b used for coding speech and television signals; b bandwidth of the information signal, n number of bits per sample ($f_b = nf_c$). In television f_c is often derived from the frequency f_{sc} of the colour subcarrier.

Information signal	System	b	f_c	n	f_b	
Speech	PCM	4 kHz	8 kHz	8	64 kbit/s	CCITT standard
TV (PAL, f_{sc} = 4.433 MHz)	PCM	5 MHz	$3f_{sc}$ = 13.30 MHz	8	106.4 Mbit/s	Experimental
	PCM	5 MHz	$2f_{sc}$ = 8.867 MHz	8	70.9 Mbit/s	
	DPCM	5 MHz	$2f_{sc}$ = 8.867 MHz	6	53.2 Mbit/s	
Video telephone	PCM	1 MHz	2 MHz	6	12 Mbit/s	
				7	14 Mbit/s	
				8	16 Mbit/s	
	DPCM	1 MHz	2 MHz	4	8 Mbit/s	

Differential pulse-code modulation (DPCM)

In audio signals, and more particularly in speech signals, the low frequencies generally predominate. This also applies to video signals: abrupt brightness transitions in the picture are the exception, a fairly even distribution of brightness is the rule. Thus, for both audio and video signals, consecutive samples often differ little in value. For this reason it may be advantageous to code not the sampling value itself but the difference between a sample and the preceding one. This is known as differential pulse-code modulation (see Table I). *Fig. 59* gives a block diagram of the equipment required. In the sampling interval k the sample A_k is compared with an approximation to the previous sample $\tilde{A}_{k-1}$ in differential amplifier D. The difference is quantized and coded (*Cod*). The *coded* signal is then transmitted, but besides being decoded in the receiver it is also decoded in the transmitter (*Decod*) and added in the integrator I to the value $\tilde{A}_{k-1}$ that was stored in I. In the next sampling period, I thus passes on to D an approximation $\tilde{A}_k$ to A_k for comparison with the new sample A_{k+1}. The receiver also consists of a decoder and an integrator.

If we now limit the difference to a region much smaller than that covered by the sampling values themselves, we can either divide up the region much more finely with the same number of levels, and hence reduce the quantization noise, or divide it up just as finely with far fewer levels, which implies a reduction in the bit rate required. The advantage gained here is paid for with a 'rounding off' of large transitions between the samples ('slope overload'), since a large change in signal level can only be covered by several steps; see *fig. 60*.

A level scheme that covers the whole region but is much more finely divided up for the small differences than for the large ones ('nonlinear DPCM', *fig. 61*) may often be a useful compromise. In television applications, for example, there is then little quantization noise in the fairly even patches of brightness in the picture, while the greater amount of noise that occurs with the larger transitions in brightness is not very annoying. The DPCM system can be refined still further by keeping the level scheme flexible and adapting it at

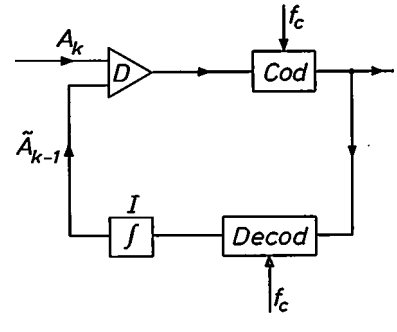


Fig. 59. Block diagram for DPCM. D differential amplifier, *Cod* coder, *Decod* decoder, I integrator, f_c sampling frequency. In the sampling interval k the difference between the sample A_k and an approximation to the preceding sample $\tilde{A}_{k-1}$ is coded. The pulse train obtained is decoded in the transmitter, and added to $\tilde{A}_{k-1}$ in I; this yields $\tilde{A}_k$ for comparison with A_{k+1} in the new interval.

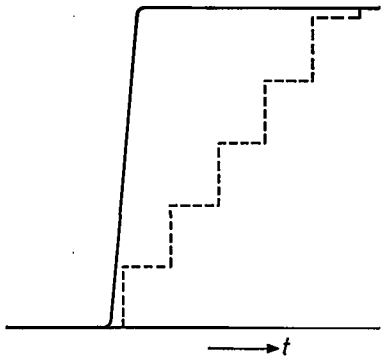


Fig. 60. In DPCM the maximum difference per step is often made considerably less than the possible difference between two samples. This permits a reduction in quantization noise. A considerable difference in consecutive samples can then only be bridged, however, in several steps, so that the transient is rounded off ('slope overload'). Solid line: original signal; dashed line: approximation after DPCM.

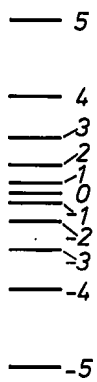


Fig. 61. Nonlinear level scheme for PDCM. Because of the fine division for the small differences the quantization noise remains low during slowly changing signals, and slope overload is avoided by the presence of large differences (coarsely divided).

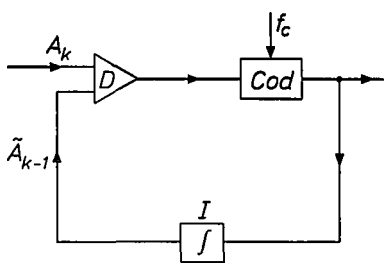


Fig. 62. Block diagram for delta modulation. The differential amplifier D determines only whether A_k is greater or smaller than $\tilde{A}_{k-1}$, which is translated by the coder as a pulse or no pulse. The integrator I converts this into a step up or a step down, as appropriate. The result is a stepped curve as shown in fig. 63.

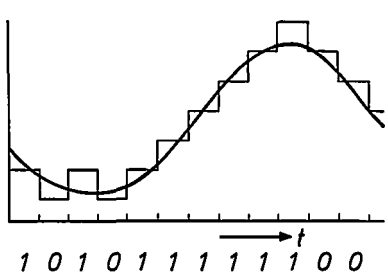


Fig. 63. Stepped curve produced from an analog signal in delta modulation.

any given moment to the situation. If, for example, a few positive steps have been made one after the other, we can then extend the level scheme to cope with the expected slope overload. This can bring considerable improvement, but does of course require more complicated equipment.

Delta modulation (DM)

A method that is remarkable for the simplicity of the equipment is delta modulation [12] (DM, *fig. 62*). This is DPCM for which only one bit per sample is used; the bit rate is equal to the sampling rate. For the sample A_k in this case it is only the sign and not the magnitude of the difference from $\tilde{A}_{k-1}$ that is determined in the differential amplifier. The coding is also simple: a pulse ('1') is transmitted when A_k is larger and no pulse ('0') when A_k is smaller than $\tilde{A}_{k-1}$. In the first case the integrator output makes a single (fixed) step upwards, and in the second case one step downwards. The result is a stepped curve that approximates to the signal (*fig. 63*). The choice of step size is a compromise here between a considerable chance of slope overload (small steps) and a high quantization noise level (large steps). The DM system, which is especially suitable for speech, has several refined variations; a system with 'digitally controlled companding' [13], for example, is used in the experimental video-telephone network in the Netherlands [14].

Effect of bit errors

Owing to noise peaks during transmission, bit errors will occasionally occur. In ordinary PCM one bit error leads to one incorrect sample value. In DPCM, on the other hand, the errors are cumulative, so that the mean level departs from the correct value.

For audio signals the mean level is of no significance; at the receiving end it is always reduced to zero. The slow accumulation of errors in DPCM and DM therefore causes no trouble in audio reception. Furthermore, an error in a difference is far less troublesome than an error in the value itself; in PCM an error in the most significant bit can be serious. For audio signals, therefore, DPCM and DM are preferable to PCM.

[12] J. F. Schouten, F. de Jager and J. A. Greefkes, Philips tech. Rev. 13, 237, 1951/52.
J. A. Greefkes and F. de Jager, Philips Res. Repts. 23, 233, 1968.
[13] J. A. Greefkes and K. Riemens, Philips tech. Rev. 31, 335, 1970.
[14] E. A. Aagaard, P. M. van den Avoort and F. W. de Vrijer, Philips tech. Rev. 36, 85, 1976.
H. P. J. Boudewijns, E. C. Dijkmans, P. W. Millenaar, N. A. M. Verhoeckx and C. H. J. Vos, Philips tech. Rev. 36, 233, 1976.

In television the situation is entirely different. While it is true that the detected signal is restored to 'black' after each line of a field, it remains a drawback that when DPCM is used an error that occurs during a line persists up to the end of the line. With DPCM, therefore, received bit errors lead to stripes in the picture, with PCM only to small flecks (*fig. 64*). The advantage of a lower bit rate or less quantization noise is therefore gained at the expense of greater susceptibility to bit errors. There are methods, however, of correcting the effect of bit errors or masking it, and DPCM is in fact often preferred to PCM for television.



Fig. 64. Television-picture degradation due to bit errors in transmission, *above:* with DPCM ($P_e = 3.3 \times 10^{-4}$), *below:* with PCM ($P_e = 10^{-2}$). The probabilities P_e of a bit error for the two cases have been chosen so as to give about the same degradation. DPCM is thus about 30 times more sensitive to bit errors than PCM.

Predictive coding

DPCM and DM can be regarded as simple cases of 'predictive coding'. The principle is to extrapolate, in accordance with a certain rule, from preceding samples

to an expected value for the sample at the time t_k. The difference between the actual value and the predicted value is coded and transmitted. At the receiver end the preceding samples and the rule are known, so that the predicted value can be determined and, with the received difference value, the actual value found from it. In DPCM and in the form of DM described above, the rule is simply that 'the predicted value is equal to the preceding sample'. Often, however, a better predicted value can be made, for example by means of linear or quadratic extrapolation from the two or three preceding samples. This is the method used in the more refined forms of DM and DPCM.

For television it is better to extrapolate in a different way. In a television picture (*fig. 65*) high correlations

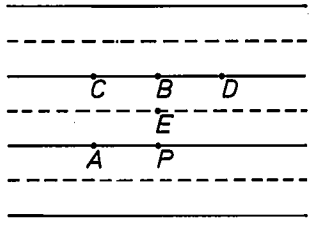


Fig. 65. Predictive coding in television. The figure represents lines and points in a small part of a television picture; the solid and broken lines represent interlaced fields. Because of the strong correlation in a picture, the luminance at P at a given instant can be reasonably accurately be predicted, for example from the signal value of the previous picture point (A) and from the values obtained one line period previously (B), immediately before and after that (C and D), one field period previously (E), and one picture period previously (P itself).

exist between a given point P in the picture and its environment (A, B, C, D, E, . . .). Since a complete picture in a picture period is generated by two successive interlaced line fields (picture period = 2× field period), it is obvious that the predicted value for the sample that will give the luminance at P at a given moment should be determined not only by the preceding sample (luminance at A) but also by the samples taken a line period before (B), immediately before and after that (C,D), a field period before (E) and possibly a complete picture period before (P itself). The predictive rule might then be: 'P shall be just as much brighter than A as B is brighter than C'. The counterpart of the advantage obtained — a reduction in bit rate — is in the first place that a memory is required for these operations that has at least the capacity of a line, a field or a complete picture, as the case may be.

It is also possible to transmit nothing at all if the actual value differs from the predicted value by less than a threshold value. Because of the high correlation between successive pictures in television, the bit rate can be substantially reduced in this way. From the resultant non-constant bit flow a strongly reduced but constant bit flow can be made with the aid of a buffer memory. This does however require additional address bits to establish the picture points to which the transmitted bits relate. This is the system used for transmitting weather charts by meteorological satellites. The method is of great advantage here, because there is very little difference between successive pictures.

Other coding methods

To conclude, two other coding methods currently of interest will be briefly mentioned.

In 'transform coding' a 'block' of samples is transformed into an equal number of other values that are subsequently coded and transmitted. The transformation may be a Fourier transform, where the coefficients of the Fourier components are transmitted, but other transformations can also be used. Two advantages are aimed at. In the first place the errors in the transmission and errors due to quantization noise will be more uniformly distributed among the original samples. In the second place, certain coefficients (e.g. those of the low-frequency or those of the high-frequency components) will often prove to be relatively insignificant, and neglecting them can lead to a reduction in the required bit rate. In television a block is often chosen in which the samples belong to a square in the picture, e.g. of 16×16 image points [15].

The other method of coding to be mentioned in conclusion offers advantages in applications such as facsimile transmission. This involves long series or runs of identical samples ('white' or 'black'). It is then preferable not to code each sample separately but to take the sample value and the length of a run. This method, known as 'run-length coding', can be improved by taking into account the correlation between successive lines.

[15] H. Bacchi and H. Tchen, Ann. Télécomm. 30, 363, 1975.

* *
*



An underground line amplifier (repeater) for 60-MHz cable transmission. The cable contains twelve coaxial cores, each carrying 10 800 telephone conversations in frequency-division multiplex (FDM). A thirteenth core with its amplifier repeater serves as a service channel. These line repeaters are spaced at intervals of $1^1/_2$ kilometres (about 1 mile).

IV. Transmission of digital signals

In the foregoing we have seen that it can be advantageous to 'package' information in pulse trains. For telecommunication it is of course essential that the pulse trains can also be transmitted. The ability to do this, however, is by no means a matter of course, the reason being that the pulse train, owing to its broad frequency spectrum, is difficult to transmit without distortion. In this part we shall discuss the measures required for the transmission of quantized pulse trains or digital signals.

If we consider the modulated pulse train itself as an information signal, then it can either be transmitted in the baseband or as the modulation of a carrier. Baseband transmission, which we shall deal with first, is used for example in digitized telephony in cases where the pulse trains obtained by PCM and TDM are transmitted by cable. Transmission on a carrier signal will be discussed in connection with data transmission, which relates to the transfer of data in digital form between locations such as the central computer of a bank or an airline company and a branch of the bank or a travel agency. Where connections are made via special cables or networks transmission is again often directly in the baseband. However, nowadays more and more use is being made of the public telephone system for such transmissions, and the digital signal then has to be transmitted as modulation of a carrier.

Baseband transmission

The situation as regards baseband transmission may be summarized briefly as follows. The distortion suffered by a pulse train (*fig. 66a*) during transmission is usually insignificant if the pulse repetition frequency (pulse rate) is much smaller than the available bandwidth. The input signal can then readily be recognized in the output signal (fig. 66*b*). When the pulse rate is gradually raised a limit is finally reached, the 'Nyquist rate', at which, although the signal is already more or less unrecognizably distorted (fig. 66*c*), it is nevertheless just possible, if certain conditions are fulfilled, to reconstruct the input signal faithfully by sampling the output signal (fig. 66*d*). A clock signal synchronizing the times of sampling to the input signal is then required. This is referred to as synchronous transmission.

Transmission at the Nyquist rate can best be illustrated by taking a hypothetical case. We consider a communication system with the amplitude characteristic of a flat lowpass filter that cuts off sharply at the frequency f_1 (*fig. 67a*), while the phase characteristic is linear (the delay is the same for all frequencies). The response at the end of this system to an infinitely narrow pulse ('δ-function') at the input is a rounded pulse, with oscillatory 'tails' (fig. 67*b*). The response has the form $A(\sin 2\pi f_1 t)/2\pi f_1 t$, if we put the zero of the time scale at the centre of the output pulse; the peak height A of the pulse is proportional to the 'content' $\int p(t)\mathrm{d}t$ of the input pulse. The transmission of a digital signal is now made possible in principle by the *equidistance of the zeros* $t = n/2f_1$ ($n = \pm 1, \pm 2, \ldots$). For if we apply to the input a train of pulses at intervals $1/2f_1$ (i.e. at the pulse rate $2f_1$), then each value at a time $m/2f_1$ (m integer) in the response is determined only by the corresponding pulse at the input; at that point in time the other pulses make no contribution. Thus, if we sample the output signal exactly at the times $m/2f_1$

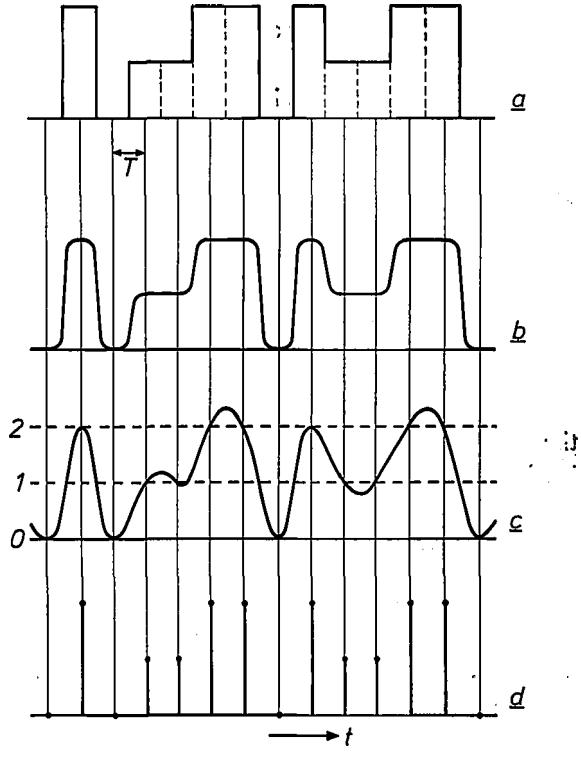


Fig. 66. *a*) Digital input signal in the form of a pulse train. T repetition period. Presentation of the pulse train in the form shown here, where each pulse value is held until the next pulse arrives, considerably reduces the part of the spectrum at the highest frequencies ('sample and hold', see also fig. 49). *b*) Output signal of a transmission system with low distortion. *c*) Output signal at the Nyquist rate. *d*) The regenerated digital signal, obtained by sampling (c) at the appropriate times.

there is no 'intersymbol interference' and we obtain a faithful reproduction of the input signal.

If now we want to transmit a signal with a pulse rate f_c less than $2f_1$, we need only an extra filter to cut off the response above $f_c/2$. The value $2f_1$ thus constitutes a maximum for the pulse rate. This maximum is the 'Nyquist rate' of the filter.

One way of understanding the 'Nyquist relation' $f_{c\,max} = 2f_1$ is to consider that, out of all the binary series that represent analog signals and are transmitted at a given elementary pulse rate, the series ...10101010... represents the signal with the highest frequency. This highest frequency is half the elementary pulse rate, and the filter must therefore pass that particular frequency, but no more is necessary. It is interesting to note that this relation is the same as that between the bandwidth of an analog signal and the repetition rate of the pulses with which the signal can be representatively sampled (p. 331).

A transmission system with the characteristics shown in fig. 67*a* is not of course a practical possibility. Good approximations to it are complicated, and have the additional disadvantage of a high group delay. This is immediately clear, since the response will closely resemble that shown in fig. 67*b*, i.e. with a long 'tail' in front of it, which is the *result*, however, of the input pulse and cannot therefore precede it. There is an important class of filters, however, that also have equidistant zeros in their response but are much easier to make. In these filters the amplitude characteristic has skew symmetry at the band ends (*fig. 68a*) and, as in the previous filter, the phase characteristic is linear. The response (fig. 68*b*) again has zeros at the times $n/2f_1$ ($n \neq 0$), where f_1 is now the skew-symmetry point of the filter curve. The less steep the 'roll-off' becomes the faster the response decreases with increasing distance from the main pulse. Crosstalk between pulses, which occurs when the sampling times are not exactly correct, is then reduced, so that the tolerance for time errors during sampling is increased. At the same time, the bandwidth required becomes larger for a given f_1, i.e. for a given pulse rate ($2f_1$).

In practice, of course, δ-pulses are never applied to the input of a transmission system, nor in general are very narrow pulses. In applying these theories we must therefore regard the pulse shaper at the input of the transmission system as a part of the system, looking upon it as a hypothetical filter that produces the actual pulses from δ-pulses.

The quality of a transmission path carrying a digital signal can be judged by means of an 'eye pattern'. This is obtained by displaying the output signal before sampling (fig. 66*c*) on an oscilloscope with a sweep rate equal to the frequency of the pulse train or of a subharmonic. If the transmission path is of good quality, patterns like those shown in *fig. 69* are observed; the signals of the successive periods are all superimposed, but at a particular instant in each period — the ideal sampling time — the signal has only one of the discrete values of the input signal. Fig. 69 shows the eye pattern of a binary signal and of a four-level

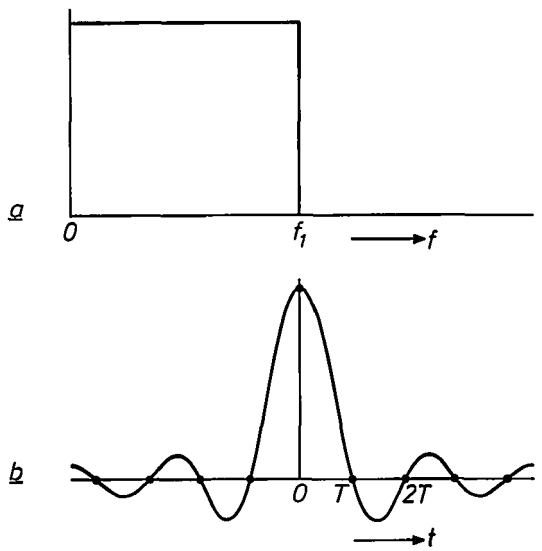


Fig. 67. A hypothetical filter with the amplitude characteristic as in (*a*) and a linear phase characteristic responds to a δ-function at the input with a signal at the output as shown in (*b*). The distance between the zeros is $T = 1/2f_1$, except at $t = 0$.

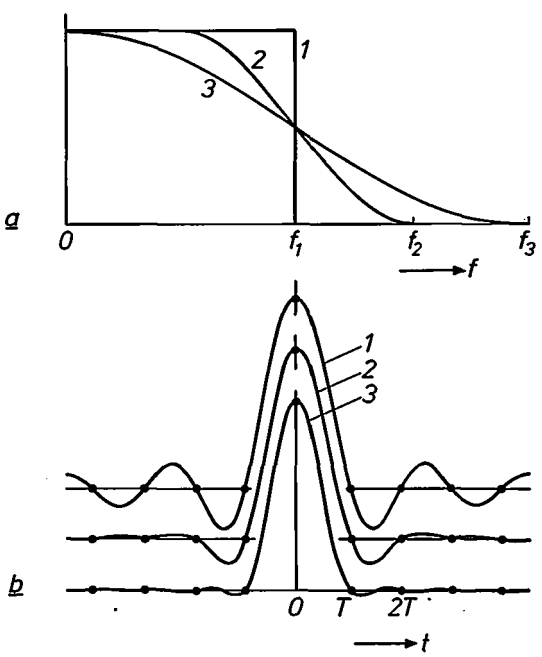


Fig. 68. Amplitude characteristics of filters (*a*) with equidistant zeros in the response (*b*) to a δ-function. The sides of the amplitude characteristics are skew-symmetric. The phase characteristics are assumed to be linear. Except at $t = 0$, the spacing between the zeros is equal to $T = 1/2f_1$, where f_1 is the skew-symmetry point of the side of the filter response. Filters are shown (*2* and *3*) whose cut-off frequencies f_2 and f_3 are 50% and 100% higher than f_1 respectively. Note that the response of *3* has other zeros besides the equidistant ones.

signal. The threshold for deciding between the levels is located in the middle of the 'eye'. The tolerances in the level determination and in the sampling time are given by the opening of the eye. At a given maximum pulse amplitude (given peak power) the eye and hence the tolerances will be smaller the larger the number of levels.

Noise in the transmission path has the effect of blurring the crossover points in the eye pattern, thereby reducing the tolerances and increasing the probability of an incorrect decision. In *fig. 70* the theoretical probability of an incorrect decision, P_e, is plotted as a function of the signal-to-noise ratio in the transmission channel. It can be seen that at an S/N of 15 dB the probability of an error with a binary system is much smaller than 10^{-6}, but with a 16-level system it is nearly 1.

Multilevel systems are therefore more susceptible to noise than a binary system. On the other hand, at a given pulse and sampling rate, the information rate — the number of bits per second — is greater. If the pulse rate (the number of 'elementary signal periods' per second) is a s^{-1}, we say we have a signal of a bauds (Bd). The bandwidth required for such a signal is at least $\frac{1}{2}a$ Hz. With a multilevel system of n bits per pulse the information rate is then na bits/s.

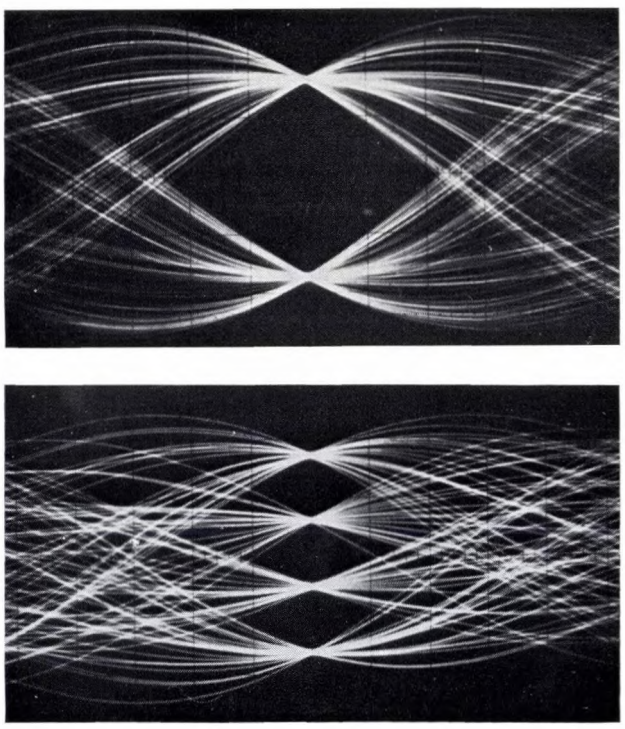


Fig. 69. 'Eye patterns' of a binary signal and a four-level signal, obtained by displaying the output signal before sampling on an oscilloscope with a sweep rate f_c (or $\frac{1}{2}f_c$, $\frac{1}{3}f_c$, ...).

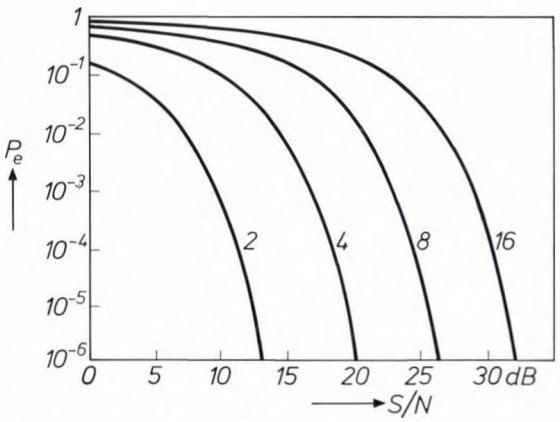


Fig. 70. Theoretical probability of an incorrect decision, P_e, as a function of the signal-to-noise ratio in the transmission channel, S/N, for systems with 2, 4, 8 or 16 levels.

At a given Nyquist rate $2f_1$, the practical filters *2* and *3* in fig. 68*a* use considerably more bandwidth (f_2 and f_3 respectively) than the 'ideal' filter *1*. However, if a *well defined symbol interference* that can subsequently be corrected is acceptable instead of 'no symbol interference', then it is possible to obtain filters for the pulse rate $2f_1$ that are *not* broader than f_1. An example is the filter with the amplitude characteristic shown in *fig. 71a* and, as usual, a linear phase characteristic; its response to a δ-function is given by fig. 71*b*. Thus, at the instants of sampling there is interference only with the immediately preceding symbol, and this interference is exactly 100%. In the case of a binary input signal the output

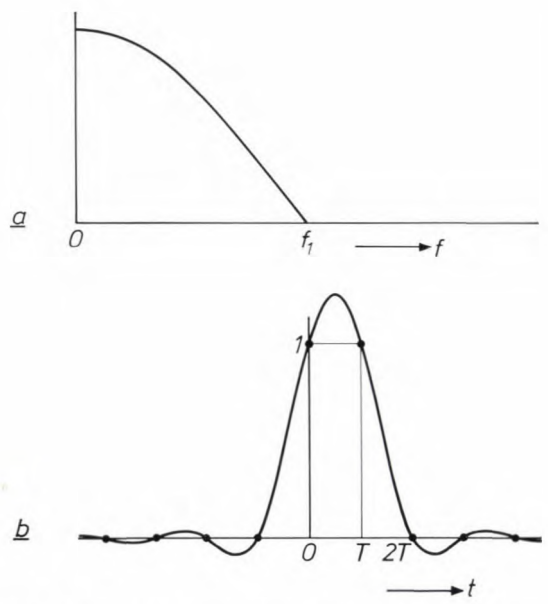
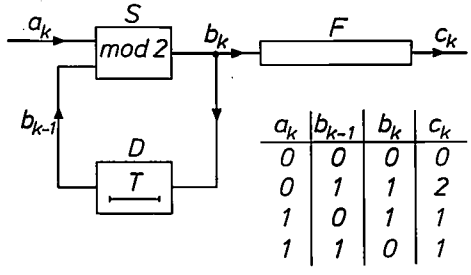


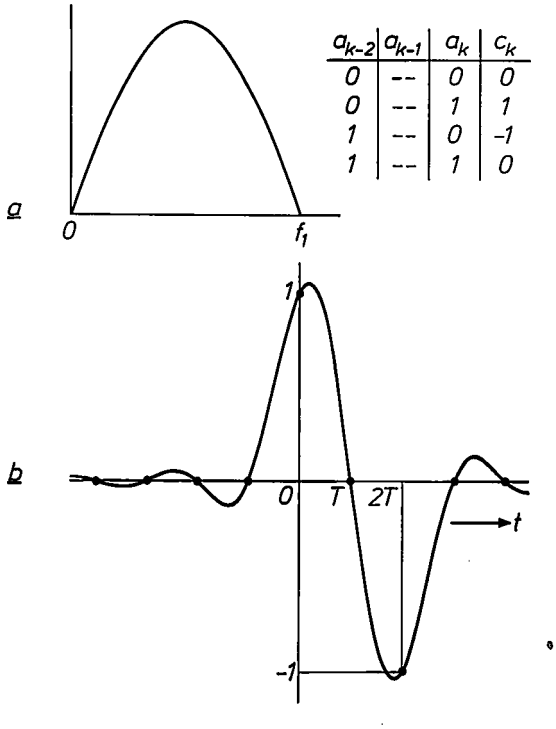
Fig. 71. Practical filter with upper limit f_1, permitting transmission with a pulse repetition rate equal to the Nyquist rate $2f_1$, but with 100% interference between two consecutive symbols. *a*) Filter amplitude characteristic. For $0 \leqslant f \leqslant f_1$ this is proportional to $\cos(\pi f/2f_1)$. The phase characteristic is linear. *b*) Response to a δ-function.

signal at the sampling times is *three-valued*; the pulse pairs (0,0), (0,1), (1,0), (1,1) at the input give at the output, at the sampling time that corresponds to the second pulse, the result 0, 1, 1, 2. If the output symbol is 0 or 2, we may conclude that the input signal is 0 or 1. If, however, it is 1, then the input symbol must be derived from the immediately preceding input symbol, which is known.



a_k	b_{k-1}	b_k	c_k
0	0	0	0
0	1	1	2
1	0	1	1
1	1	0	1

Fig. 72. Recoding of the input signal to avoid error propagation, when the filter of fig. 71 is used. In S the recoded symbol b_k is formed from the input symbol a_k and the recoded symbol of the preceding pulse repetition period, b_{k-1}, by the addition in modulo 2. (The result of the 'modulo-2 addition' is the last digit of the sum in the binary system; in particular, for 1 and 1 the result is 0.) D single-period delay line. F transmission system with the characteristic of fig. 71. The table gives the response of the system to the various possible combinations of a_k and b_{k-1}. The input signal follows unambiguously from the output symbol, which was not the case without recoding.



a_{k-2}	a_{k-1}	a_k	c_k
0	--	0	0
0	--	1	1
1	--	0	-1
1	--	1	0

Fig. 73. Filter characteristic that can be obtained, unlike the previous filters, with many existing communication systems, because of the filter transmission 0 at frequency 0. The figure gives the amplitude characteristic — sin $(\pi f/f_1)$ for $0 \leqslant f \leqslant f_1$; the phase characteristic is again linear — the response to a δ-function ($T = 1/2f_1$), and the response table for transmission at the Nyquist rate $2f_1$.

If a series of ones appears at the output of this system, and one of these is the result of a transmission error, that error will be propagated in the interpretation of the input symbols right up to the end of the series. This can be avoided by recoding the input signal. Before transmission the input series a_k is converted by the circuit shown in *fig. 72* into the series b_k, which follows from a_k and b_{k-1} in accordance with the given response table. It can be seen that we can now unambiguously decide on the value a_k from the value of the output symbol c_k.

The advantage of having the highest possible pulse rate at the given bandwidth is paid for here by a higher susceptibility to interference and noise, since the output signal is no longer binary but 'ternary'. It is found that the signal-to-noise ratio of the transmission channel must be 2.1 dB higher for the same error probability. Since the information rate is no greater than that of the binary input signal, the signal is referred to as 'pseudo-ternary' [16].

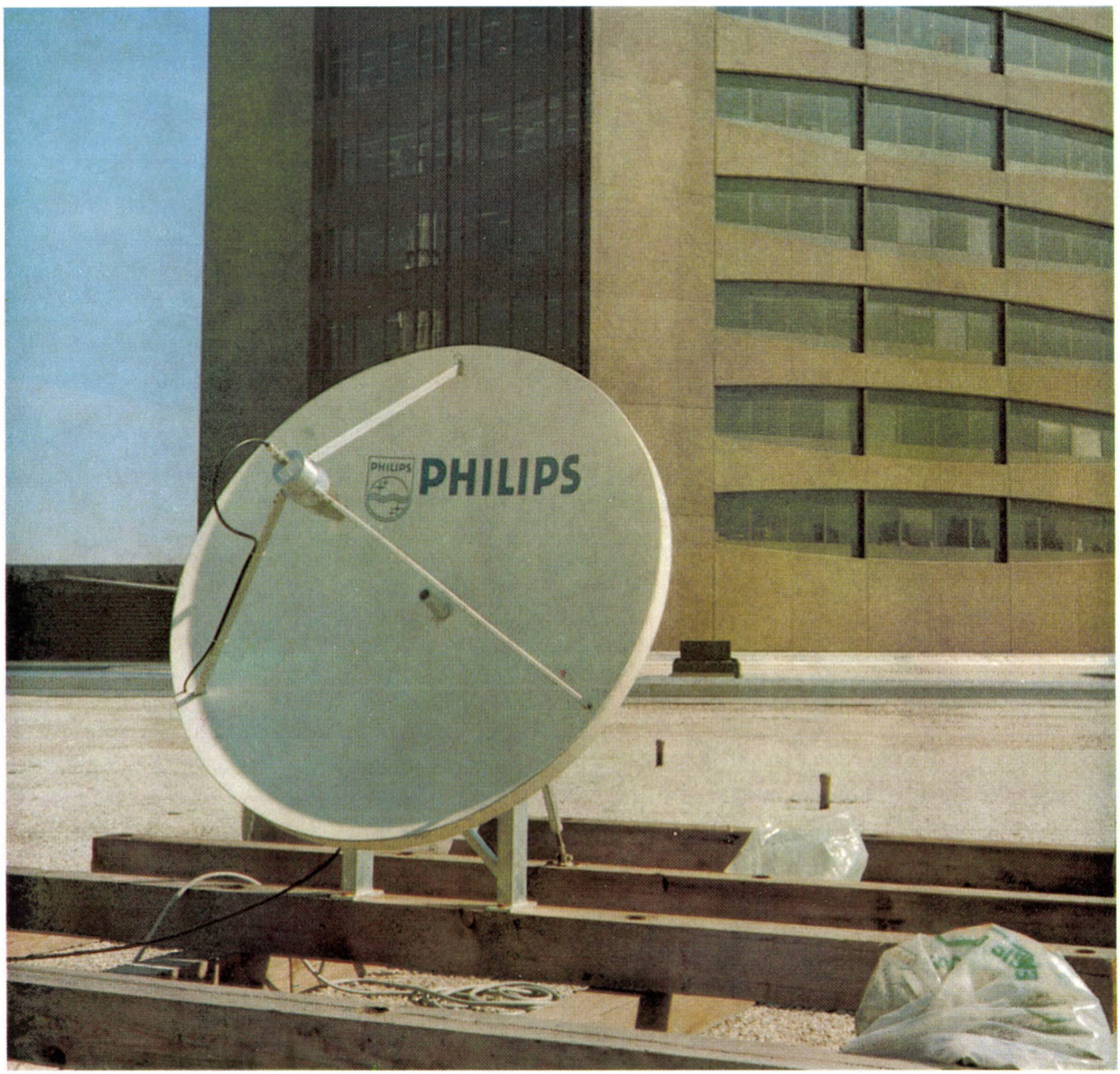
Systems with no d.c. components

The amplitude characteristic of the systems discussed above has a relatively high value at zero frequency, which implies that d.c. components are readily transmitted. In many existing circuits, however, such as cable transmission systems with isolating transformers, the d.c. component is not transmitted, and therefore the models discussed cannot be used.

In such cases the filter with the amplitude characteristic given in *fig. 73* (again with a linear phase characteristic) is a more suitable model. Here again, the passband is no wider than half the sampling rate, there is well defined symbol interference (—1 with the second preceding symbol), and the output signal is pseudo-ternary. Zero frequency is not transmitted, however: the output signal averages zero (1 occurs just as often as —1).

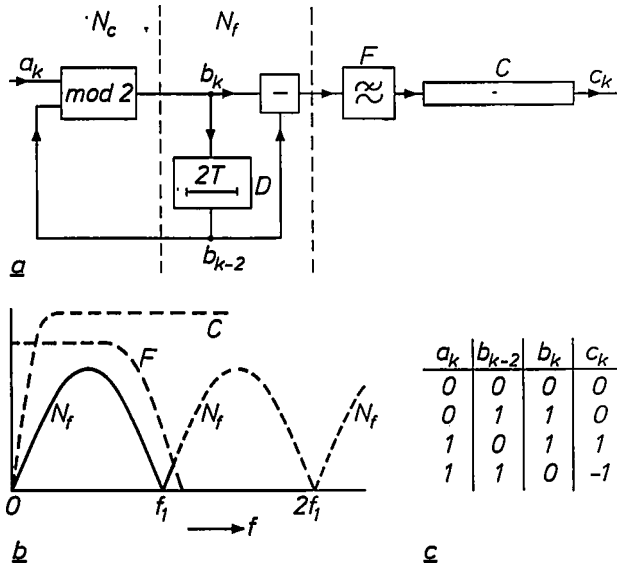
Fig. 74a shows schematically how this method is often realized in practice. We shall disregard for the moment the section N_c. In N_f the signal delayed by two pulse periods, b_{k-2}, is subtracted from the input signal b_k. The result is that signals with frequencies 0, $\frac{1}{2}f_c, f_c, \frac{3}{2}f_c$ . . . are not transmitted; the characteristic is thus as shown in fig. 74*b* (curve N_f). The filter F cuts off everything except the first lobe of the response, resulting in the characteristic of fig. 73. The original transmission system (C), e.g. a cable system with isolating transformers, has a fairly flat response, except for the sharp drop at $f = 0$, and therefore the charac-

[16] P. J. van Gerwen, Philips Res. Repts. **20**, 469, 1965.





Reception from television satellite with parabolic antenna. The photograph shows a trial installation on the roof of the Maison de Radio Canada, Montreal, Quebec. The diameter of the antenna is 1.20 m (about 4 feet). The reception frequency is about 12 GHz, the modulation FM. The installation is suitable for private use; an FM/AM converter is then fitted to the television receiver (see photograph on the left). Several sound channels can be combined in one video channel, making it possible to receive the same programme in different languages on different frequency channels.



a_k	b_{k-2}	b_k	c_k
0	0	0	0
0	1	1	0
1	0	1	1
1	1	0	-1

c

Fig. 74. *a*) Arrangement for obtaining the amplitude characteristics in fig. 73. *b*) Response of sections N_f, F and C. *c*) Response table. The response N_f in (*b*) is obtained by subtracting b_{k-2} from b_k; the delay line D gives a delay of two pulse periods. Because of the filter F, only the first lobe of the response of N_f in (*b*) is left, and the transmission system C changes this situation very little. The section N_c (modulo-2 addition of a_k and b_{k-2}) gives suitable recoding for preventing error propagation.

Clock extraction

In the systems discussed it is essential that the sampling should occur at the correct times on reception. The receiver therefore usually contains a 'clock-pulse generator', which delivers pulses at the appropriate sampling times. The rhythm of this 'clock' can be derived from the received signal ('clock extraction'). An example of a synchronizing circuit is shown schematically in *fig. 75*. For correct clock extraction the received signal should contain sufficient transitions between 0 and 1. A 'scrambler' is often used for this; it changes the input pulse train in accordance with a fixed rule such that the original pulse train can be reconstituted after reception. Some coding methods

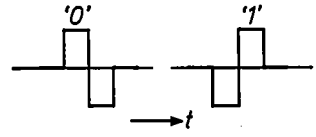


Fig. 76. Biphase coding. The advantages of this coding are: 1) good clock extraction, because each symbol gives a transition, 2) the signal has no d.c. component. The bandwidth required is at least twice the Nyquist bandwidth, however.

teristic undergoes little change. The section N_c serves for recoding the input signal a_k in such a way as to prevent the propagation of errors.

As can be seen from the response table, the circuit has the useful feature that a simple rectifier alone is all that is necessary to reconstitute the input signal at the output after sampling. The circuit described is a form of *bipolar coding* (of the second order, so called because of the delay by two pulse periods in fig. 74; 'first-order bipolar coding' is also widely used).

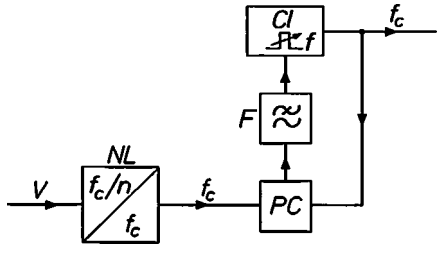


Fig. 75. Clock extraction. If f_c is the elementary pulse repetition frequency, the output signal V of the transmission system does not usually contain the frequency f_c itself but subharmonics of it (frequencies f_c/n, n integer) or various components that differ in frequency by f_c or f_c/n. A signal of frequency f_c synchronous with V is then obtained with the nonlinear element *NL*. Depending on the coding system used, *NL* simply has to rectify the signal or carry out a more complicated nonlinear operation. With the aid of the phase comparator *PC* and the lowpass filter F, the phase of this signal establishes the frequency and phase of the clock signal produced by the clock generator *Cl*. *PC*, F and *Cl* form a 'phase-locked loop'.

result automatically in many transitions; in this respect *biphase coding* is particularly favourable. *Fig. 76* gives the elementary signals that this method uses for 0 and 1. It can be seen that there is a transition in each elementary period. (This method of coding also has no d.c. component. The pulse repetition rate, however, is no more than half the Nyquist rate.)

A synchronized clock at the receiving end is characteristic of what are known as 'synchronous systems'. Older slow systems such as telegraphy and telex often operate asynchronously. If sampling is required, each letter and each numeral (coded in a number of bits) is preceded by a start signal for the generator in the receiver that delivers the sampling pulses. This generator is required to remain in step for only a small number of bits, so that further control is unnecessary.

Equalization

In the systems discussed so far it is equally essential that the whole system between the transmitter and the receiver of the digital signal (including the hypothetical filter mentioned above, representing the pulse shaper) should in fact have the selected amplitude and phase characteristic.

If the data transmission takes place in the baseband, in a system in which a connection is established in a different way each time in a switching centre, the characteristics of the transmission path are variable and not well known. At the receiving end the characteristic

must then be 'equalized' by means of additional, variable filters [17]. The same applies for data transmission with modulation via a long-distance telephone network, which will presently be discussed.

Applications

An application of the baseband transmission of digital signals has just been mentioned: data transmission in special data networks. Another application, also mentioned earlier, is found in digitized telephony. When the digital TDM signal is transferred between two main exchanges by cable, it is sent in baseband. Coding ensures that the d.c. component is zero and that the number of transitions is always sufficiently large for good clock extraction. The CCITT has standardized four bit rates: 2, 8, 34 and 140 Mbit/s. Four signals of 2 Mbit/s can be grouped in TDM to form a signal of 8 Mbit/s, etc. At the higher bit rates some latitude is left in the TDM grouping for additional signals.

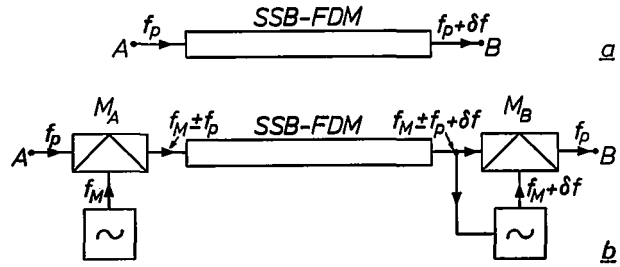


Fig. 77. *a*) The Fourier components of a signal in the speech band, after transmission over an SSB-FDM telephone system, are in general shifted in frequency by a small amount δf. This is no great disadvantage for speech, but for digital signals it is unacceptable. *b*) Transmission of digital signals by telephone by modulation of a carrier (frequency f_M) in the speech band. Modulation and demodulation take place in 'modems' (M_A and M_B) at the subscriber's terminals. In data transmission from A to B, M_A acts as a modulator and M_B as a demodulator, and vice versa. Since the reference signal for the demodulation is derived from the output signal, it receives the same frequency shift. This shift is therefore eliminated in the demodulated signal.

Modulation systems for digital signals; data transmission

The transmission of signals between digitized telephone exchanges over a microwave link is a practical example of signal transmission on a carrier rather than in the baseband. Since a good signal-to-noise ratio is extremely important and bandwidth presents no problems, frequency modulation is used. We shall not pursue this example further but will confine ourselves to data transmission via the telephone system.

When a digital signal in the baseband is applied to one end of a long-distance transmission path, the chance that it will arrive at the other end in a recognizable form is remote. In the first place the amplitude characteristic — a speech band from 300 to 3400 Hz — does not extend to zero frequency, and as we saw above, this introduces distortion. (The above analysis also applies in the situation considered here, and in fact to any linear system.) What is more, the phase characteristic is strongly curved. Now it might if necessary be possible to overcome these difficulties by means of coding and equalization. An insurmountable difficulty, however, is that the baseband spectrum is almost always shifted in frequency by a small amount, δf say, in the transmission path — in general an FDM grouping of SSB signals — as a result of the SSB modulation and demodulation; see *fig. 77a*. This means that the fixed frequency and phase relations between the original Fourier components of a signal are destroyed. For example, a periodic rectangular signal at the input will arrive at the output not only deformed but also having lost its periodicity. If we try to obtain an eye pattern at the output, we see a constantly changing picture in which at the very best an 'eye' will appear only every now and then.

The answer is modulation of a carrier in the speech band (fig. 77*b*). A 'modem' (modulator-demodulator) is connected to each of the two subscriber's terminals. When data signals are sent from A to B the modem at A acts as a modulator and modulates the digital signal on a carrier of frequency f_M (e.g. 1800 Hz). The modem at B acts as a demodulator. The carrier for demodulation is extracted from the signal before the modem, and thus receives the same frequency shift δf. After demodulation the frequency difference has therefore been eliminated. At the same time there is now a much freer choice of transmission characteristic; the telephone network will accept any characteristic provided it is not too wide for the speech band into which it is transposed by the modulation.

In the early days of data transmission the data signals consisted of no more than a few hundred bauds. Frequency modulation was adopted for the sake of simplicity. This method has the added advantage of providing a good protection from the strong variations in transmitted amplitude on the telephone network. The method used was asychronous. Frequency modulation here was called frequency-shift keying (FSK), a term originating from telegraphy.

Systems of this type are still used for signals up to 1200 Bd. In the system for 1200 Bd the frequency jumps between 1300 Hz ('1') and 2100 Hz ('0'). This can be regarded as frequency modulation of a carrier at 1700 Hz with a deviation of 400 Hz. The baseband width given by Nyquist's rule is at least 600 Hz and the bandwidth required for the FSK channel is found on

[17] A. Gersho, Bell Syst. tech. J. 48, 55, 1969.
F. de Jager and M. Christiaens, to appear in Philips tech. Rev. 37, No. 1, 1977.

applying Carson's rule to be 2000 Hz. The FSK band (700-2700 Hz) is easily accommodated in the speech channel. The modulation index is small (400/600 = 0.67).

The bit rate for telegraph and telex signals is lower. FDM in a speech channel is used when such signals have to be transmitted over the telephone network. It is usual to subdivide a speech channel into 24 120-Hz channels for bit rates up to 50 bit/s, or twelve 240-Hz channels for bit rates up to 100 bit/s. By using a 1200-Bd system in one speech channel, it would be possible to achieve the same information rate by TDM grouping of the original signals (1200 bit/s = 12 channels of 100 bit/s = 24 channels of 50 bit/s), but this involves considerable practical complications.

In recent years a growing need has arisen for higher bit rates. These require synchronous operation in the first place to achieve the required sampling accuracy. Furthermore an FSK signal of 2400 bauds with a frequency deviation of ± 400 Hz no longer fits into the speech channel, and a smaller frequency deviation soon gives rise to excessive noise. More generally, FSK uses up a considerable bandwidth; this is closely related to the fact that it is a nonlinear modulation method.

In this respect 'phase modulation' (PSK, phase-shift keying) is better. In PSK the '0', for example, is represented by a phase shift of 0°, the '1' by a phase shift of 180° with respect to the nominal phase of the carrier (two-phase PSK, *fig. 78*). This form of phase modulation is not essentially different from amplitude modulation in which the amplitude changes sign, and may also be regarded as a linear system. If the system is well designed, the bandwidth required at a given pulse rate is not much greater than the Nyquist bandwidth. A signal of 2400 Bd — at least 1200 Hz wide — could in fact be accommodated in a speech channel.

In practice the speech channels are often of poor quality. To safeguard against this, systems with more than two phase states are used, making it possible to transmit a larger number of bits per second with signals of a given number of bauds (i.e. of a given bandwidth). At each moment of sampling the phase may be 0°, 90°,

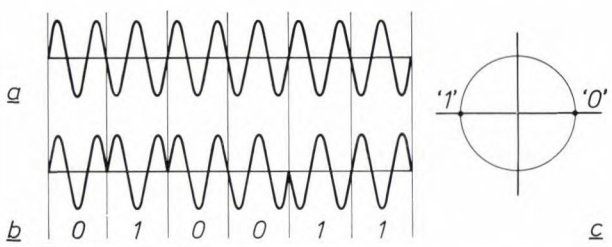


Fig. 78. Two-phase PSK ('phase-shift keying'). *a*) Carrier. *b*) The modulated signal for a digital signal 010011. In the case illustrated, carrier and digital signal are synchronized. This is not always the case in practice. *c*) The 'signal-constellation points' '0' and '1' in the vector diagram.

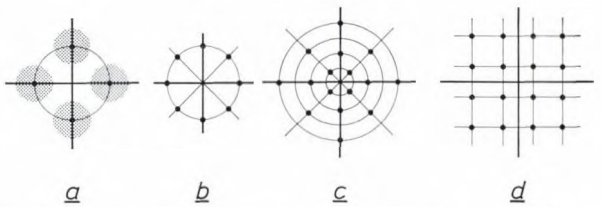
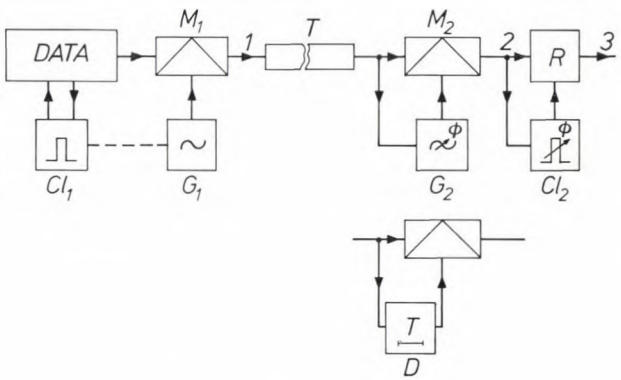


Fig. 79. 'Signal constellations' in the vector diagram for the various forms of PSK under experiment or already in use: PSK with four phases (*a*), eight phases (*b*), four amplitudes and four phases per amplitude (*c*) and the system with 'four-level quadrature modulation' (*d*). Noise circles are shown only in (*a*). The closer the points lie together, the greater the probability of errors due to noise. In (*d*) there are just as many signal-constellation points as in (*c*). Case (*d*) gives better noise performance because the points are better distributed, but (*c*) lends itself better to the compensation of errors due to 'phase jitter'.

180° or 270° (four-phase PSK, *fig. 79a*). Since the phase states in the vector diagram lie closer together, the 'noise circles' touch one another earlier; four-phase PSK is consequently 3 dB more sensitive to noise than two-phase PSK.

This development has been extended, leading for example to systems with eight phases, systems with four amplitudes and four phases per amplitude, and



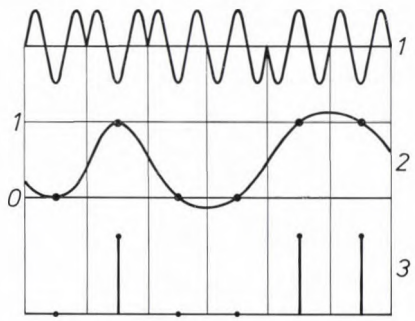


Fig. 80. Synchronous PSK data-transmission system. *T* telephone connection, M_1, M_2 modems. G_1 carrier-signal oscillator, G_2 reference oscillator. Cl_1 data clock, Cl_2 sampling clock, *R* data-signal regenerator. *1*, *2* and *3* (schematic) are the signals after M_1, between M_2 and *R* and after *R*. The synchronization for the demodulation in M_2 and the synchronization for the regeneration in *R* are in principle distinct from each other, but in some cases Cl_1 and G_1 are also synchronized (dashed line; see fig. 78). To avoid the complications of 'carrier extraction' from the signal before M_2 or of pilot signals, this signal is often compared not with a regenerated carrier signal but with the signal from one pulse repetition period before (*D* delay element for one pulse repetition period).

systems with 'four-level QAM', which go up to 9600 bit/s (see fig. 79*b*, *c*, *d* and *Table II*).

Compared with modems for 'slow' FSK the modems for 'fast' PSK have two refinements: carrier synchronization for the demodulation of the received signal, and bit synchronization for the regeneration of the digital signal from the demodulated signal; see *fig. 80*. The second feature requires 'clock extraction'. For synchronous demodulation it is necessary to have a reference signal, which can be obtained from a separate pilot signal. To avoid this complication, the phase is often compared not with the phase of the original carrier but with the phase in the previous pulse period (differential PSK; fig. 80 with the second version of the modem M_2). The reference here is therefore the preceding phase state, but since this is also affected by noise, the signal-to-noise ratio with this simplification is worse (2.3 dB in four-phase PSK, 2.8 dB in eight-phase PSK).

In general the higher the bit rate the more sensitive the system is to interference and noise. Data transmission at the higher bit rates is therefore only possible if the telephony channel is not too unsatisfactory. Variations in the transmission will always have to be equalized, preferably automatically. In FDM telephone channels trouble may be experienced from the noise effect known as 'phase jitter', caused by instabilities in the oscillators of the FDM system.

To meet the need for ever higher bit rates, systems for more than 9600 bit/s have been developed. It is difficult to accommodate these in a single speech channel. One of the solutions to this problem is to give the subscriber a direct line to a main exchange, where he is given the use of an entire basic group (60-108 kHz), which normally comprises 12 speech channels. CCITT standards for this also exist (Table II). To economize on bandwidth, SSB or VSB modulation is used. There are systems, for instance, for 48, 56, 64 and 72 kbit/s with lower sideband modulation on a carrier of 100 kHz, using pseudo-ternary coding.

Finally, we should note that the required signal processing, such as filtering, modulation and clock extraction, is being increasingly digitized in modern modems [18]. This dispenses with the need for large inductors and capacitors, simplifies the design and adjustment of the system and allows the use of integrated circuits.

Table II. Modem systems standardized by the CCITT.

Transmission over one speech channel (300-3400 Hz)

Bit rate (bit/s)	Pulse freq. (Bd)	Method	Carrier * (Hz)	sync./ async. **	CCITT code
200-300	200-300	FSK	980-1180 / 1650-1850	async. / async.	V21
600	600	FSK	1300-1700	async.	V23
1200	1200	FSK	1300-2100	async.	V23
2400	1200	PSK, 4 phases	1800	sync.	V26
4800	1600	PSK, 8 phases	1800	sync.	V27
9600	2400	PSK, 4 amp. ×4 phases	1700	sync.	V29

Transmission via an FDM basic group (60-108 Hz)

(kbit/s)			(kHz)		
40.8		VSB	100	async.	V35
48-56-64-72		LSB	100	sync.	V36

* In FSK the two frequencies are given here that represent the two bits.
** Sampling of the output signal synchronized (sync.) or not synchronized (async.) with the input signal.

[18] P. J. van Gerwen, Philips tech. Rev. 30, 71, 1969.
F. A. M. Snijders, N. A. M. Verhoeckx, H. A. van Essen and P. J. van Gerwen, IEEE Trans. COM-23, 1259, 1975.

* *
*



Adjustment of a 60-MHz modulator for carrier telephony. The 60-MHz carrier can be modulated by 10 800 speech channels in frequency-division multiplex (FDM); single-sideband modulation is used.

V. Modulation in telecommunication

Modulation and telecommunication are closely interrelated. Apart from the simplest wire or cable connections for telephony and telegraphy, telecommunication is unthinkable without modulation. It is also true to say that all known modulation systems were really developed for telecommunication purposes. In this part we shall therefore give a survey of the specific uses of modulation in the principal forms of telecommunication in current use. Other areas of application touched on earlier were the storage of information (FM in magnetic recording of video signals, PFM and PDM with the 'VLP' record) and signal processing (PAM with charge-transfer devices). There are other areas where modulation is used, but not as such an essential feature: range-finding and navigation systems for traffic on land, at sea and in the air, remote control and telemetry, but we shall not be concerned with these here.

The principal forms of telecommunication are shown in *fig. 81*. There are four groups: broadcasting, telephony, fixed links and mobile systems. Under 'broadcasting' we shall presently discuss a few more modulation aspects of AM and FM sound broadcasting, television broadcasting and broadcasting via satellites. We shall not return here to telephony as such; the modulation methods used in telephony — frequency-division multiplexing of SSB signals and time-division multiplexing of pulse-coded signals — have already been dealt with at some length. We shall, however, take a closer look at the 'fixed links' between two stations. These are not unrelated to broadcasting and telephony but are integral links in these systems, for example between studio and transmitter in broadcasting, and between two exchanges in telephony. Fixed *radio* links operate either at frequencies below 30 MHz or in the UHF and SHF bands (microwaves); they are

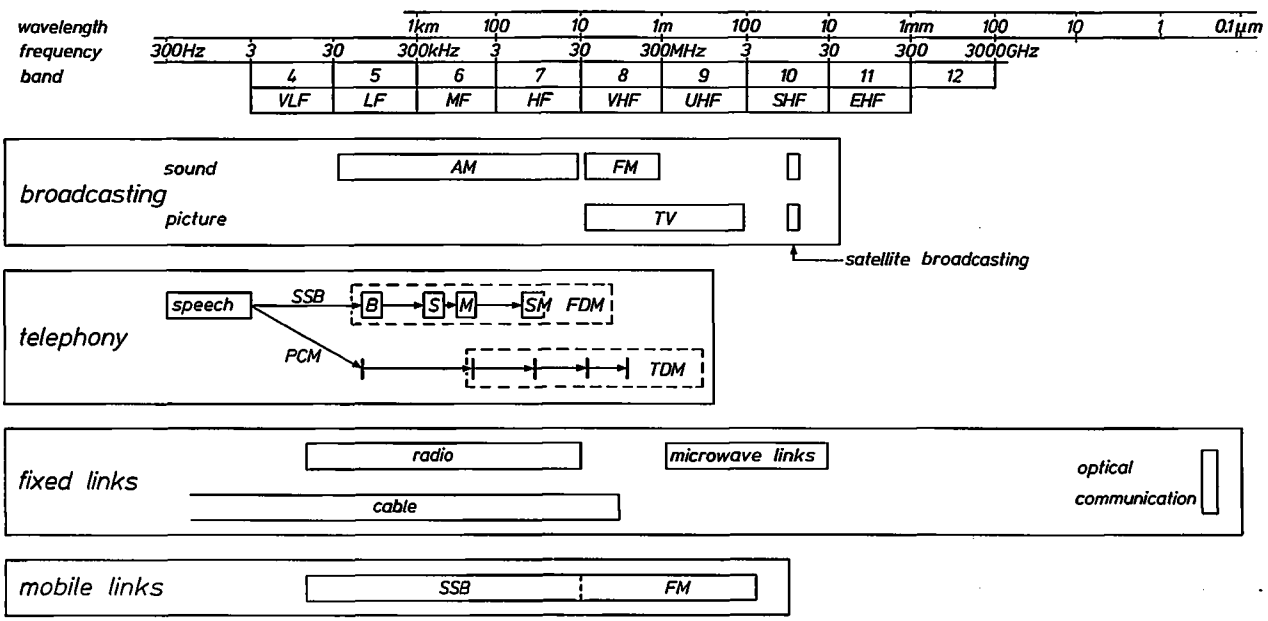


Fig. 81. Forms of telecommunication. *Broadcasting*: AM radio in bands *5*, *6* and *7*; FM radio in band *8*, television in bands *8* and *9*. Satellite broadcasting — with direct transmission from satellite to individual receivers — is intended for television and sound broadcasting (band *10*). Band *n* is the band from 0.3 to 3 times 10^n Hz; the frequency 10^n Hz thus lies at the centre of this band. *Telephony*: frequency-division multiplexing of speech signals to form 'basic, super, master and supermaster groups', and time-division multiplexing of signals coded in the form of pulses. *Fixed links*: transmission paths between two fixed stations for broadcasting and telephony (and also for telegraphy, telex, data transmission, ...). *Cable* links are, very generally, used for short distances but are also in use for very large distances, especially in telephony (transoceanic cables). *Radio* links below 30 MHz are slightly directional and have a relatively long range; those in bands *9* and *10* (microwave links) are strongly directional and have a relatively short range (about 50 km), unless they include communication satellites. In the latest form of *optical communication* a type of cable link is again used for guiding the electromagnetic waves (here light waves), the cables in this case being glass fibres. *Mobile systems* are used for communications with and between pedestrians, cars, ships, aircraft, etc. For frequencies below 30 MHz these generally use SSB modulation, but at higher frequencies FM is preferred, although AM is also still in use.

only weakly directional in the first case and highly directional in the second. The microwave links have been given a new dimension by communication satellites. Very much higher up in the spectrum we have optical communication, which in its modern form is still in the experimental stage. Finally we shall discuss systems where at least one of the stations is mobile.

Before going further with this, two comments should be made on the subject of transmission by cable, but we shall leave the matter at that. In conventional broadcasting the transmission generally takes place in free space, but the systems that have a single central antenna for small residential estates, suburbs, or towns ('relay television') form an exception. In relay systems the 'transmitter' is a relay station that receives the signals from elsewhere through a large antenna or in some other way. The subscriber to such a system generally has a wider choice of programmes, and pictures and sound of better quality, than is possible with individual reception. To permit the use of the same receivers, the signal is generally similar to the conventionally transmitted signals, and for that reason these systems need not be considered here. The old form of 'rediffusion system', in which an audio sgnal was distributed in the baseband by cable, has now almost disappeared from the scene.

'Fixed links' are often cable systems. For small-scale communications, for example within a group of buildings, this is nearly always the case. For telephony in particular there are also cable links that are tens, hundreds or thousands of kilometres long (transoceanic cables), over which FDM and TDM groups are transmitted (the TDM groups in the baseband, see p. 349). The exponential attenuation of the signals is compensated by regularly spaced repeaters.

Broadcasting

AM sound broadcasting

For sound broadcasting in bands 5, 6 and 7, the familiar 'long-wave', 'medium-wave' and 'short-wave' bands, the modulation system used is conventional AM. The radio signals in this wavelength range are able to propagate over great distances by bending and reflection from the ionosphere, so that one transmitter of sufficient power covers a large area. To avoid interference between the large number of transmissions the frequency spectrum is rigorously divided up into many narrow bands, and the audio bandwidth allocated for many transmitters is limited to 5 kHz, which means that each transmitter takes up 10 kHz of the r.f. spectrum.

It was long argued that the replacement of conventional AM by SSB modulation would allow broadcasters to use twice as much audio bandwidth, making it possible either to increase the quality of the transmissions or to increase the number of transmitters. The old counter-argument that an SSB receiver is much more expensive than a receiver using simple peak detection is much weakened by the continued advances in electronics (ICs, LSI). There remains, however, the fact that there are still large numbers of conventional receivers in use, and these receive SSB signals severely distorted. As a transitional measure the use of special SSB systems has been proposed, where the carrier would not be completely suppressed and with the sidebands pre-distorted in such a way as to compensate for the distortion caused in the conventional receiver ('compatible single sideband', CSSB) [19]. This system, however, causes distortion in a real SSB receiver. The whole question is still being studied by the CCIR.

FM sound broadcasting

For FM sound broadcasting in band 8 a much wider bandwidth is available. In the first place, band 8 (like every other band) is nine times as wide in linear frequency measure as all the lower bands put together, and it is linear measure that counts for audio bandwidth. In the second place, these shorter waves undergo less bending and they are not reflected by the ionosphere, which means that the range of the transmitters is shorter and that two transmitters a few hundred kilometres apart can have the same frequency without interfering with one another. Advantage is taken of this frequency latitude to broadcast FM transmissions of high sound quality. In general each transmitter is allocated an r.f. bandwidth of 200 kHz. In this bandwidth a broad audio spectrum (15 kHz) can then be combined with a frequency deviation of say 75 kHz, thus giving strong noise suppression (see (16) and p. 323).

We shall touch briefly on some extensions and experiments for which there was and still is room in FM sound broadcasting, and which depend on modulated subcarriers. In the first place the system has been extended to include stereo sound [20]. In conformity with a recommendation made by the CCIR, the signal $S(t)$ of a stereo FM transmitter is made up as shown in *fig. 82a*. From the 'left-hand audio signal' L and the 'right-hand audio signal' R the sum and difference are formed. A subcarrier at 38 kHz is modulated in DSB, with suppressed carrier, by the difference signal. The information signal $p(t)$ for the transmitter modulator is the sum of the sum signal $L + R$, the modulated subcarrier and a synchronous pilot signal at 19 kHz; fig. 82*b* shows the spectrum of $p(t)$. The pilot signal serves for the synchronous demodulation of the modulated subcarrier in the FM receiver. It is also some-

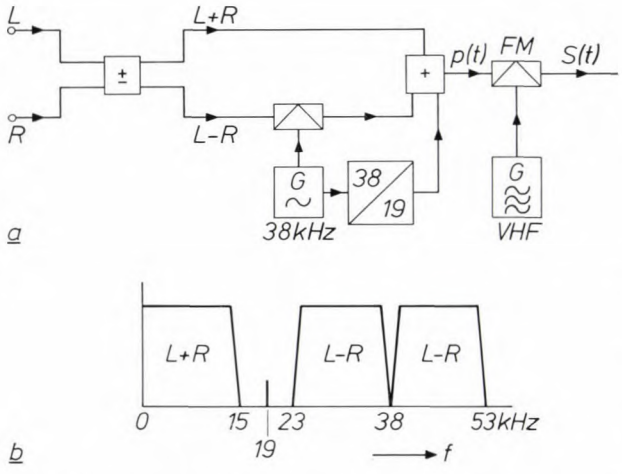


Fig. 82. Stereophony in FM radio. *a*) Diagram showing the composition of the FM signal $S(t)$. L left-hand audio signal, R right-hand audio signal. The difference signal $L - R$ DSB-modulates a (suppressed) subcarrier at 38 kHz. *b*) Spectrum of the information signal $p(t)$ for the frequency modulator FM; $p(t)$ is the sum of the sum signal $L + R$, the modulated subcarrier and a pilot signal of 19 kHz synchronized with this carrier. The pilot signal serves as a reference during the reconstitution of $L - R$ from the subcarrier in the FM receiver.

times used in stereo transmissions to activate the $(L - R)$ channel in the receiver; without a pilot signal this channel is closed in these receivers. The system gives compatability with mono receivers, which only receive the sum signal $L + R$. At a frequency deviation of 75 kHz the required bandwidth is now about 260 kHz ($2\times(53 + 75) = 256$). The total frequency deviation in this system is not much greater than for a mono transmitter, which, for the same L and R, transmits only $L + R$ (*fig. 83*). The signal-to-noise ratio in the $(L + R)$ channel may thus be about as good as in the mono transmitter. In the $(L - R)$ channel, however, f_{max} is about three times higher, and the modulation index is thus three times smaller, so that the gain in signal-to-noise ratio is about ten times less than in the $(L + R)$ channel. For stereo, therefore, the coverage of an FM transmitter is smaller than for mono. Stereo reception will therefore often require a good outside antenna where for mono reception a simple inside antenna is sufficient.

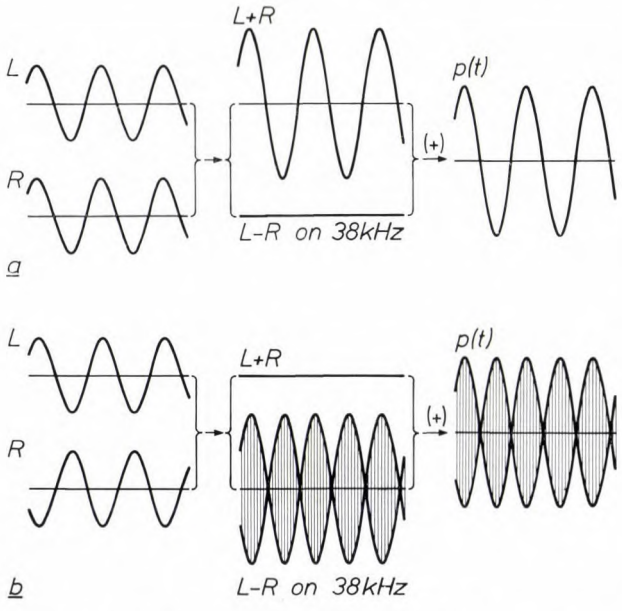


Fig. 83. The information signal $p(t)$ in fig. 82 for cases where L and R are sinusoidal signals of identical amplitude and are in phase (*a*) or in antiphase (*b*). The vertical strips in (*b*) represent the subcarrier. In *both* cases the amplitude of $p(t)$, and hence the frequency deviation of $S(t)$ in fig. 82, have the same magnitude as for a mono transmitter with $2R$ or $2L$ as the information signal. It can be shown that the same applies approximately for all other combinations of L and R.

The idea of using a single FM transmitter to broadcast extra signals on subcarriers in addition to the main signal is currently finding application in various ways. In the United States some FM transmitters broadcast a continuous programme of background music that frequency-modulates a subcarrier at 67 kHz (sometimes 41 kHz in the case of mono transmitters); the listener can choose between the main programme or the background music. In the Netherlands the Dutch broadcasting organization and Philips are carrying out joint experiments in which FSK (65-67 kHz) is used for sending digital information together with a broadcast programme; the digital information is then displayed on a panel in the receiver. In Germany some transmitters send a digital signal, modulating a carrier at 57 kHz, to indicate when traffic information is to be given. In a car radio tuned to the transmitter this signal then switches the programme on again which the driver had switched off. A problem with systems of this kind is the possible crosstalk that such signals can produce on the $(L + R)$ and $(L - R)$ signal, as a consequence of nonlinearities and delay errors in the receiver. A subcarrier of 66 kHz, for example, can interfere with the pilot signal at 19 kHz to give a signal of 47 kHz, which produces a 9-kHz tone in the $(L - R)$ channel. Introduction of the system therefore calls for some caution, and this matter is also being studied by the CCIR.

Television broadcasting

Television broadcasts are found in bands 8 and 9 (see fig. 81). It was decided at the time, for economy of bandwidth and the possibility of peak detection, to adopt vestigial-sideband amplitude modulation (VSB-AM, see p. 317) for transmitting the video signal. In practice the transmitter sends out the conventionally modulated carrier (carrier plus two sidebands), with

[19] Th. J. van Kessel, F. L. H. M. Stumpers and J. M. A. Uyen, E.B.U. Rev. No. 71A, 12, 1962.

[20] See for example N. van Hurck, F. L. H. M. Stumpers and M. Weeda, Philips tech. Rev. **26**, 327, 1965 and **27**, 62, 1966.



Semitrans 102 modem, acoustically coupled to a telephone. This modem is used for feeding digital signals into the public telephone system. The two levels of the binary signal are coded as two audio frequencies (frequency-shift keying or FSK); these are transmitted by the microphone and are supplied to the modem at the receiving end via an acoustic connection to the telephone. Two pairs of frequencies are used — 980/1180 Hz and 1650/1850 Hz — so that simultaneous two-way communication is possible; the maximum transmission rate is 200 bits per second.

the lower sideband almost completely suppressed (*fig. 84*). The filter with the symmetrical response characteristic of VSB modulation (the 'Nyquist characteristic') is in the receiver. The sound modulates (AM in France, FM elsewhere) a separate carrier, which is 6.0 MHz above the video carrier in the system used in Britain. For colour transmissions the colour information is carried on a subcarrier in the video band. The system for modulating the subcarrier depends on which colour-television system is used (QAM for NTSC and PAL, FM for SECAM) [8].

In television broadcasting, experiments designed to transmit more information by means of additional subcarriers are also in progress. In particular, trials are being carried out with a second sound channel. This could for example be used for a second language when an event of international importance is being reported. In Germany this is done with a second sound carrier 242 kHz above the first. In Japanese trials the 'second sound' frequency-modulates a subcarrier at about 32 kHz, which, together with the 'first-sound', frequency-modulates the sound carrier ('FM-FM system')

Satellite broadcasting

In addition to broadcasting by 'Earth' stations, we now have broadcasting by satellite, i.e. by a transmitter in a space satellite that covers a part of the Earth where the signals can be received by individual sets. Signals at 12 GHz in band 10 are now being used experimentally for such transmissions. The problem of sufficient signal strength is much greater here than in the case of a satellite for a 'fixed link'. In a fixed link the Earth station can use a very large parabolic antenna; for individual reception of satellite broadcasts, on the other hand, a parabolic 'dish' of about 1 metre in diameter would seem to be about the maximum. For a satellite in the geostationary orbit (see p. 359) to cover a country like Germany, France or Great Britain with a television programme, an 80-kW transmitter would be necessary if the conventional AM-VSB system was used. This would not be a practical proposition in a satellite. If FM was used, however, with adequate bandwidth, then 500 W would be sufficient; this power is feasible in a satellite. In the CCIR a system has been proposed that would use FM with a deviation of 8 MHz in a 27-MHz band. Efforts are now being made to reach international agreement on the use of the 12-GHz band for television and sound broadcasting.

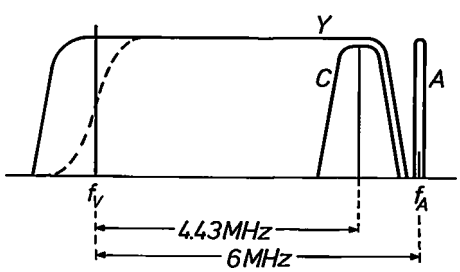


Fig. 84. Spectrum of a television signal. In principle, vestigial-sideband amplitude modulation (VSB-AM, p. 317) is used for transmitting the video signal (carrier frequency f_V). The transmitter does not, however, transmit the actual VSB signal but a conventional AM signal most of whose lower sideband is suppressed. A filter in the receiver, preceding the actual detector, provides the skew-symmetric VSB response (dashed line). The video signal consists of the luminance signal Y and the chrominance signal C (a subcarrier at 4.43 MHz modulated by the colour signals $R-Y$ and $B-Y$). Special measures are needed to ensure that the effect of C on Y and vice versa does not give picture distortion [8]. The sound usually modulates a separate carrier at a frequency f_A (in Britain $f_A - f_V$ is equal to 6 MHz).

Carson's rule (16) gives the required bandwidth of 27 MHz directly from the deviation of 8 MHz and the video bandwidth of 5.5 MHz. The fact that this combinaton also gives the desired improvement in S/N is partly because the eye is much more sensitive to noise at the lower frequencies than at the higher frequencies in the television picture. This is accounted for in the S/N calculations by means of a weighting factor; this is to the advantage of FM, which of course gives less noise at the lower video frequencies than at the higher ones ('triangular noise', p. 322). To simplify a little, the eye is only sensitive to noise at frequencies up to 1 MHz in the video signal. With $f_{max} = 1$ MHz and $\Delta f = 8$ MHz we find $\alpha = 8$, and for the improvement in S/N compared with AM we thus find a factor of $3 \times 8^2 \approx 190$ (see p. 323), which broadly agrees with the factor of 160 between the figures mentioned above of 80 kW for AM and 500 W for FM.

Fixed links

The conventional 'fixed radio links' — the connection formed by radio signals at frequencies up to about 30 MHz between two fixed stations — used to be the principal means employed in telecommunication (telephony, telegraphy, telex) over long distances. The contact, which depends on the signal being reflected from the ionosphere, is not in general very reliable because of the varying and to some extent unpredictable degree of ionization in the various layers of the ionosphere. Many such links have therefore been replaced by cables or satellite links. Where they are still used, the optimum modulation systems have generally been introduced, i.e. SSB and ISB (see p. 313 and 316). Formerly conventional AM was usual.

Microwave links

Radio signals at frequencies in bands 9 and 10 (wavelengths from 1 cm to 1 m) can readily be beamed with the aid of parabolic reflectors, and the microwave beams thus obtained nowadays play an important part as links in transmission paths for FM radio, television and telephony. Because of beam bending and diffraction, the range of the quasi-optical beam is in principle

somewhat farther than the horizon. With masts about 50 m high, ranges of about 50 km can be covered. Repeater stations are used to cover greater distances. The scattering of microwaves by the troposphere also makes direct contact possible between two stations several hundreds of kilometres apart; this method is often employed for military communication.

Frequency modulation is the obvious method to use here, since power is more of a problem than bandwidth. At the transmitter the power of the microwave beam is no more than 20 W, and is often much less. For two microwave links more than a few hundred km apart the same frequency can be used. Furthermore, FM provides the signal that is least vulnerable to the nonlinearity of the amplifiers and the considerable amplitude fluctuations that often occur (fading due to fluctuating propagation conditions).

Fig. 85 shows a conventional modulation system and a possible distribution of the successive signals over the spectrum. The frequencies and bandwidths quoted are practical examples. The information signal frequency-modulates a carrier at 70 MHz in a bandwidth of 30 MHz. The 30-MHz band is then transposed to 3884 MHz by SSB modulation of a (suppressed) carrier at 3954 MHz; this is known as 'up-conversion'. This stepwise modulation is necessary because direct frequency modulation of carriers in the GHz band

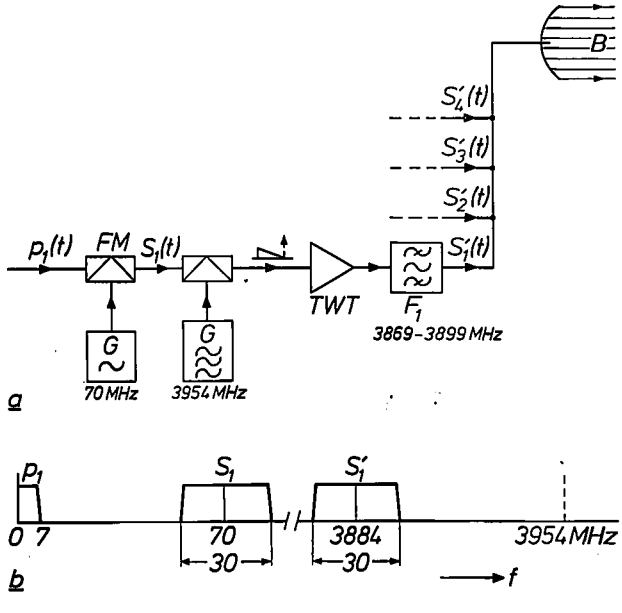


Fig. 85. Transmission of television signals and multiplexed telephone signals over a microwave link. *a*) Modulation, *b*) spectra of the various signals. The information signal $p_1(t)$ frequency-modulates a 70-MHz carrier with a bandwidth of 30 MHz (*FM*). The resulting signal $S_1(t)$ is transposed to a microwave frequency by lower-sideband modulation ('up-conversion'). *TWT* travelling-wave tube. F_1 filter that passes only the lower sideband. S_1' the signal after F_1. The microwave beam B contains microwave signals S_1', S_2', S_3', . . . , which are obtained with different SHF carrier frequencies and are spaced some distance apart from each other in the spectrum.

— obtained by multiplication from crystal oscillations — is not a practical proposition. Signals obtained in this way are frequency-division multiplexed on a single microwave beam, with the GHz carrier frequencies spaced by about 30 MHz or a multiple of this value. The '4-GHz band' (3800-4200 MHz) was the first to be reserved for fixed microwave radio links. There is ample room in this band for 2×12 channels of 30 MHz, some of which are used as service channels. The factor of 2 is due to the use of horizontally as well as vertically polarized beams. When the beams are used for telephony, the signal for each direction is sent via a different channel. At a given repeater station the frequencies of the incoming beams must differ from those of the outgoing beams. It is also necessary to have careful allocation of the frequency channels among the various beams in each region of a few hundred kilometers in diameter to avoid mutual interference. In many countries the 4-GHz band is becoming crowded. For microwave radio links higher-frequency bands have already been allocated and some are now in use.

Nowadays a substantial proportion of trunk calls and international calls are made via microwave radio links. The telephone signals are multiplexed by frequency division on channels of bandwidth say 30 MHz. In television, microwave beams are used for transmitting the signals from studio to transmission mast, and also, for example, in 'Eurovision' transmissions, for distributing international programmes over Western Europe. For 625-line television a frequency deviation of $\pm$ 8 MHz is often used; this leaves room for a television signal with a baseband width of nearly 8 MHz in the 30-MHz band. Prior to the main modulation the sound is made to frequency-modulate a subcarrier at a frequency of typically 7.5 MHz, added to the video signal. For international links systems are used with four sound channels on subcarriers of 7.020, 7.500, 8.065 and 8.590 MHz to transmit speech in four languages simultaneously.

During a microwave transit the signal accumulates the 'triangular noise' peculiar to frequency modulation. To distribute this noise more uniformly over the speech channels in telephony, pre-emphasis is applied to the frequency-division-multiplexed group before modulation, with an attenuation of 4 dB at the lowest frequency up to an amplification of 5 dB at the highest.

Pre-emphasis is also applied to television signals (from -11 dB at the lowest frequency to $+3$ dB at the highest), but not so much for noise reduction as to limit certain colour defects in the picture. This may be briefly illustrated with the aid of *fig. 86a*. This shows the video signal as a function of time during part of a line period when it passes through three colour areas. It consists of a luminance signal (dashed) with the colour subcarrier superimposed on it; the phase and amplitude of the subcarrier determine the colour. Colour errors easily occur in this situation because, owing to nonlinearities in the microwave-link equipment, the phase and amplitude of the colour signal are affected by the level of the

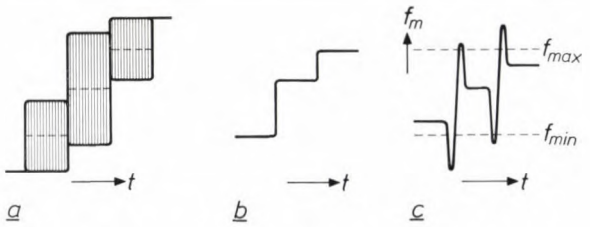


Fig. 86. *a*) Video signal during part of a line period while passing through three different colour areas. The dashed line indicates the luminance signal, with the chrominance signal superimposed on it. As a result of nonlinearities in the equipment, the level of the luminance signal influences the phase and amplitude of the chrominance signal. Pre-emphasis gives a relative reduction of the differences in levels and thus reduces the errors. Sharp transitions in the luminance (*b*), however, may lead to overshoot of the instantaneous frequency of the frequency-modulated signal if the pre-emphasis is too great (*c*).

luminance signal ('differential phase and amplification errors') [8]. These errors can be reduced by a relative reduction of the 'level differences' before frequency modulation. This is just what pre-emphasis does, since the level variations represent the lowest frequencies in the signal. Strong pre-emphasis is not permissible here, however, because if the high frequencies are over-emphasized overshoot will occur at an abrupt change in the luminance signal (sharp transition between light and dark); see fig. 86*b*,*c*. The instantaneous frequency of the VHF or UHF signal may then go outside the passband, giving rise to annoying distortion and noise.

Communication satellites

To enable them to function as 'repeater stations' in a fixed microwave link, communication satellites are put into the geostationary orbit, i.e. an orbit directly above the equator at a height of about 35 900 km (about 2.8 Earth diameters). At this distance the orbital period is equal to the Earth's period of rotation. Given the right orbital direction, the satellite will thus be stationary with respect to the Earth's surface. To help compensate for the attenuation (beam spread) due to the great distance, parabolic antennas with diameters from 10 to 25 m are used at the Earth stations. Great distances can be bridged with one satellite because more than a third of the Earth's surface is 'visible' from the satellite. The Earth transmitter sends a directed beam to the satellite, which then retransmits the received signal back to the Earth at another carrier frequency. Here again, the transmitted signals are mainly television signals and FDM groups of (900 or 1800) telephone channels.

A modern communication satellite has various repeaters ('transponders') with antennas that can be oriented ('pointed') separately. The Intelsat IV type of satellite — the first version was launched in 1971 and is in orbit above the Atlantic — has twelve transponders, each with a bandwidth of 36 MHz, two antennas that cover the whole visible part of the Earth, called 'global antennas', and two more closely beamed antennas, called 'spot' antennas. The spot antennas are used for the busiest routes. The global antenna serves for the less busy routes. Two Earth stations in the area of coverage, communicating through a global antenna, keep a complete transponder busy during the period of communication.

An exception is the transponder reserved for 'SPADE' (*S*ingle Channel *P*er Carrier Multiple *A*ccess *D*emand Assignment *E*quipment); this operates in an Intelsat IV above the Atlantic. The name indicates that each of the 800 channels of bandwidth 45 kHz into which its band of 36 MHz is divided may be temporarily used on demand by any Earth station for a connection that is not so frequently required. SPADE is also an exception in the modulation method used. The signals are coded in PCM of 64 kbit/s and then transmitted by four-phase PSK in a bandwidth of 38 kHz so that each channel retains 7 kHz as a safety margin ('guard bands'). A separate information channel is used for requesting and assigning SPADE channels.

The fact that the velocity of light is not infinitely high is beginning to cause some difficulties in telephone conversations by satellite. In a station-to-station route over one satellite the transit time of the signals is as long as a quarter of a second. For this reason it has been agreed that a telephone channel shall never include more than one satellite.

Optical communication

Telecommunication by modulating a beam of light is in itself very old; an example is the naval signalling lamp. Modern techniques allow the use of higher modulation frequencies. It is now more than ten years since a television signal was first transmitted on a light beam over a distance of a few kilometres. The light source used — an LED or a laser — was intensity-modulated and the signal was detected by a photomultiplier or a solid-state photodetector. This method had some inherent drawbacks and was not therefore taken up. Today, however, a promising development is taking place in which communication is established by the modulation of light conducted through glass fibres [21]. The potential advantages of this system, compared with the coaxial cable, include a very large bandwidth, small dimensions, low losses and insensitivity to electrical interference.

It is possible to transmit an analog signal along glass fibres, but since nonlinearities in the transducers are difficult to avoid the obvious method is to use binary coded signals and to modulate the light source by the resultant bit runs (on = '1', off = '0'). The bandwidth,

[21] See the issue of Philips Technical Review on this subject: Vol. **36**, No. 7, 1976 (pp. 177-216).

and hence the bit rate, are limited by transit-time dispersion of the light in the fibre. This is mainly attributable to the fact that the light travels along different paths through the fibre. In some types of fibre (the 'single-mode fibre' and the 'graded-index fibre') this effect is absent or greatly reduced, and bit frequencies of gigabits per second are possible.

Mobile systems

We have already seen that SSB modulation is greatly to be preferred to conventional AM for voice communication by radio between mobile units, such as ships, aircraft and automobiles, or between such a unit and a fixed station. With SSB modulation the required bandwidth is twice as small and the required transmitting power some 13 dB lower. For frequencies below 30 MHz conventional AM has therefore been almost completely superseded by SSB modulation.

Above 30 MHz frequency modulation is now widely used for mobile communication systems. Amplitude variations are of course unavoidable and very large in mobile installations, but with FM the information is well protected from the effects of such variations. Another advantage is that different FM transmitters at relatively short distances apart can use the same channel, since there is only a small range of field-strength ratios in which they can interfere with one another at the individual receivers (p. 324). This is an important consideration, for example, in mobile telephone systems. A commonly used bandwidth per channel is 25 kHz. For speech (baseband width 3.4 kHz) the frequency deviation may then be as much as 8 kHz, so that in addition there is an improvement in signal-to-noise ratio compared with AM (modulation index at least 2).

However, the advantages of SSB are also leading to the growing use of SSB modulation for frequencies above 30 MHz. The necessary stability in the local oscillators in the transmitter and in the receiver can be obtained even for mobile equipment and at higher frequencies by the use of recent techniques.



Ground station of the Netherlands Post Office at Burum providing a link with communication satellites. In the traffic via the 'Intelsat' satellite, orbiting at 35 900 kilometres above the Atlantic, certain frequency channels are allocated to the Netherlands at certain times for telephone communications with North America. The up-link uses frequencies of about 6 GHz, the down-link frequencies of about 4 GHz. Frequency modulation is employed, and the telephone signals are grouped in the conventional manner by frequency-division multiplexing (FDM).

MODULATION

INTRODUCTION 307

I. MODULATION OF A SINUSOIDAL CARRIER 309
Conventional amplitude modulation 309
Double-sideband modulation with suppressed carrier (DSB) 311
DSB modulation compared with conventional AM 312
Single-sideband modulation (SSB) 313
Generation of an SSB signal 315
Frequency-division multiplex (FDM) 316
Related methods of amplitude modulation 316
Frequency and phase modulation 318
Modulation and detection 319
Bandwidth 321
Signal-to-noise ratio 322
Pre-emphasis 323
Threshold effect 324
Applications; magnetic recording of video signals 326

II. MODULATION OF PULSE TRAINS 329
Some basic concepts 329
Pulse-amplitude modulation (PAM) 330
Time-division multiplex (TDM) 330
Nyquist's theorem 331
Pulse-frequency and pulse-duration modulation (PFM and PDM); the 'VLP' record 332

III. QUANTIZATION AND CODING OF ANALOG SIGNALS . . 337
Pulse-code modulation (PCM) 337
Differential pulse-code modulation (DPCM) 338
Delta modulation (DM) 339
Effect of bit errors 339
Predictive coding 340
Other coding methods 341

IV. TRANSMISSION OF DIGITAL SIGNALS 343
Baseband transmission 343
Systems with no d.c. components 346
Clock extraction 348
Equalization 348
Applications 349
Modulation systems for digital signals; data transmission 349

V. MODULATION IN TELECOMMUNICATION 353
Broadcasting 354
AM sound broadcasting 354
FM sound broadcasting 354
Television broadcasting 355
Satellite broadcasting 357
Fixed links 357
Microwave links 357
Communication satellites 359
Optical communication 359
Mobile systems 360